

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 2, 2003, 23:22:25 ; Search time 2898 Seconds

(without alignments)
10373.774 Million cell updates/sec

Title: US-10-027-450-46

Perfect score: 1033
Sequence: 1 atcatggcgcgcgcctctctc.....aaaaaaaaaaaaaaaaa 1033Scoring table: IDENTITY-NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl.*

1: gb.ba:*
2: gb.htg:*
3: gb.in:*
4: gb.om:*
5: gb.ov:*
6: gb.pat:*
7: gb.ph:*
8: gb.pl:*
9: gb.pr:*
10: gb.ro:*
11: gb.sts:*
12: gb.sy:*
13: gb.un:*
14: gb.vi:*
15: em.ba:*
16: em.fun:*
17: em.hum:*
18: em.in:*
19: em.mu:*
20: em.om:*
21: em.or:*
22: em.ov:*
23: em.pat:*
24: em.ph:*
25: em.pl:*
26: em.ro:*
27: em.sts:*
28: em.un:*
29: em.vi:*
30: em.htg.hum:*
31: em.htg.inv:*
32: em.htg.other:*
33: em.htg.mus:*
34: em.htg.pln:*
35: em.htg.tod:*
36: em.htg.mam:*
37: em.htg.vtl:*
38: em.sy:*
39: em.htgo.hum:*
40: em.htgo.mus:*
41: em.htgo.other:*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	462.2	44.7	131955	AP004676	AP004676 Oryza sat
2	462.2	44.7	170586	AP005006	AP005006 Oryza sat
3	460.6	44.6	339972	OSA307662	AJ0307662 Oryza sat
4	210.8	20.4	787	AY063029	AY063029 Arabidops
5	210.8	20.4	970	AY035158	AY035158 Arabidops
6	210.8	20.4	45714	AC006224	AC006224 Arabidops
7	210.8	20.4	90283	AC004450	AC004450 Arabidops
8	209.2	20.3	945	AY087084	AY087084 Arabidops
9	197.8	19.1	802	AY117208	AY117208 Arabidops
10	197.8	19.1	966	AY063828	AY063828 Arabidops
11	197.8	19.1	1014	AT060594	AT060594 Arabidops
12	196.6	19.0	975	AY065366	AY065366 Arabidops
13	196.6	19.0	89517	ATF17J16	ATF17J16 Arabidops
14	159.8	15.5	10029	AE012835	AE012835 Chlorobiu
15	77.6	7.5	9830	AFACBD	I36817 Alcaligenes
16	66.6	6.4	14034	AE005694	AE005694 Caulobact
17	65.2	6.3	4347	AB017109	AB017109 Thermus t
18	65.2	6.3	4347	E36126	E36126 L-Lysine bl
19	64	6.2	8217	AF416776	AF416776 Methyloba
20	62.4	6.0	3057	MMWDEG2	Y14634 Mus musculu
21	61.2	5.9	171187	AC116960	AC116960 Dictyoste
22	59.8	5.8	1542	AY070834	AY070834 Drosophil
23	59.6	5.8	2661	USMLEUA	L20832 Ustilago ma
24	59	5.7	97683	AC116548	AC116548 Dictyoste
25	58.6	5.7	42210	SC1C2	AL031124 Streptomy
26	58.6	5.7	58930	AC098321	AC098321 Rattus no
27	58.4	5.7	113193	AF357202	AF357202 Streptomy
28	57.8	5.6	13968	AE000901	AE000901 Methanoba
29	57.6	5.6	12454	AE002019	AE002019 Deinococc
30	57.6	5.6	12992	AE010370	AE010370 Methanopy
31	57.6	5.6	268147	AC116966	AC116966 Dictyoste
32	57.2	5.5	885	BC002080	BC002080 Mus muscu
33	57.2	5.5	1047	BC006773	BC006773 Mus muscu
34	57	5.5	1466	AR208521	AR208521 Sequence
35	57	5.5	1466	AR208522	AR208522 Sequence
36	57	5.5	1472	AR174965	AR174965 Sequence
37	57	5.5	1472	AR174966	AR174966 Sequence
38	57	5.5	1875	AR208551	AR208551 Sequence
39	57	5.5	1875	AR208552	AR208552 Sequence
40	57	5.5	1881	AF008300	AF008300 Dicotilar
41	57	5.5	1881	AR174995	AR174995 Sequence
42	57	5.5	1881	AR174996	AR174996 Sequence
43	56.8	5.5	125020	AF429315	AF429315 Homo sapl
44	56.8	5.5	595	AX090442	AX090442 Sequence
45	56.4	5.5	15348	AE007163	AE007163 Mycobacte

ALIGNMENTS

RESULT 1
AP004676/c 131955 bp DNA linear HTG 21-MAR-2002
LOCUS AP004676
DEFINITION Oryza sativa (japonica cultivar-group) chromosome 2 clone
OUT1003.B06, *** SEQUENCING IN PROGRESS ***, in ordered pieces.
ACCESSION AP004676
VERSION AP004676.1 GI:18447935
KEYWORDS HTG; HTGS; PHASE2.
SOURCE Oryza sativa (japonica cultivar-group) (cultivar: Nipponbare) DNA,
clone: OJ1003.B06.
ORGANISM Oryza sativa (japonica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzaceae; Oryza.

REFERENCE
1

[illegible]

QY	558	CGTGGGCGCTGGAAAGATGCAAGACAGGGGATGTGTACCGTGGACCTGCAATC	617
Db	87057	CACGTGAGCGCTGGAAAGATGCAAGACCGGGGATGTGTACCGTGGACCTGATATTTG	86998
QY	618	CGTTTTATTAACACACCTCTGCGAAGAGATACAACTGAACCAATTGGTGTGCTGG	677
Db	86997	CGTCATGATTAACCAACATCCGCGAAGACGATACAACTGAAGCTATGGCGCATCCCG	86938
QY	678	CCCTGTATTTGAGGCGGAGGAGATCTTTGCTTAAGCCCGGAAGACAGGATGTGTCGTC	737
Db	86937	GCCCGTTATTGAGAGCAGGCGGATCTTTGCTTAATGCCGGAAGACCGGATGTATGCCATC	86878
QY	738	GAAGACTGC	746
Db	86877	CAAGTCTGC	86869

AP005006/c	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE
AP005006	170586 bp	DNA	linear	HTG 28-MAR-2007		
Oriza sativa (japonica cultivar-group)	chromosome 2	clone P0519E06				
*** SEQUENCING IN PROGRESS ***, in ordered pieces.						
AP005006						
AP005006.1	GI:19773546					
HTG, HTGS_PHASE2.						
Oriza sativa (japonica cultivar-group)	(cultivar:Nipponbare)	DNA,				

ORGANISM	REFERENCE
<i>Oryza sativa</i> (japonica cultivar-group)	1
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta	
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;	
Ehharctoideae; Oryzaceae; <i>Oryza</i> .	

SAKAKI, T., MATSUMOTO, T., KANONAKI, T., and OIRYA, S. (2002) A genomic DNA, chromosome 2, PAC clone: F03192605 Published Only in Database (2002) 2 (bases 1 to 110586) SAKAKI, T., MATSUMOTO, T., and YAMAMOTO, K. Direct Submission Submitted (27-Mar-2002) Takuji Sasaki, National Institute of Research Program, Kanonaki

2-1-2, Tsukuba, Ibaraki 305-8602, Japan
(E-mail: tsasakiahi@affrc.go.jp, URL: <http://79p.dna.affrc.go.jp/>,
Tel:81-298-38-7441, Fax:81-298-38-7468)
NOTE: It currently consists of 1 config. Gaps between the configs
of the same version of the pieces is believed

are based on estimates as soon as it is available and the accession number will be preserved.

* NOTE: This is a 'working draft' sequence.

* This sequence will be replaced

* by the finished sequence as soon as it is available and

* the accession number will be preserved.

Location/Qualifiers

```

SOURCE
2. /organism="Oryza sativa (japonica cultivar-group)"
   /cultivar="Nippohare"
   /db_xref="taxon:39947"
   /db_xref="taxon:39947"
   /chromosome=2
   /clone="P0519606"
BASE COUNT 46926 a 37577 g 38026 g 47905 t 152 others
ORIGIN

```

Query Match	44.7%	Score 462.2	DB 2	Length 170586
Best Local Similarity	78.2%	Pred. No. 5.7e-85		
Matches 570	Conservative	1	Mismatches 149	Indels 9
				Gaps 1
QY	27	GGCGGTCCAGGCAAGCGCTTTAGAGCCCAATCCGAGTCCCAACAGGCGGTTATCCG	86	
Db	30068	GGCGGCGCCGATAGCAACAATTCTTGAGCCGTGTCCACGCCCTTCAGAGACGTTCCGCCG	30009	

/product="putative protein"
 /protein_id="CAC39031.1"
 /db_xref="GI:14140114"
 /translation="MTVAARGRREGAPARRMHGKGTITWMMRCSTYSVVA
 VLDLGEAGSGSTGRSRPSGSGGRMGMEPKLGPSGAGAGMGWRGPSNDGEGARSEE
 RIRSLQAGSSSPFLTARYCSPLCLPLLRPRVAAAVR"
 8619..8711
 /gene="W30ERIPDM"
 /number=1
 8712..8792
 /gene="W30ERIPDM"
 /number=1
 8793..9131
 /gene="W30ERIPDM"
 /number=2
 12306..12944
 /gene="W50ERIPDM"
 12306..12944
 /gene="W50ERIPDM"
 /codon_start=1
 /product="putative protein"
 /protein_id="CAC39032.1"
 /db_xref="GI:14140115"
 /translation="MERSERRERRRYAALAVILLVFASAPVRCGDADGEDGACA
 AEPCTGTCLPDPWKGSGSGSQSSPDGAMRYGWMMAASPGKSGFGFYGSGREG
 GGGGGGSGGGRAYFGGCGYGGHPPGGGSGGSGGRNKGSGGIGCYGMYGGY
 NGPEPGGGGGGEGGGYGGDYGGDVGAMNSKRSFPGGGGKTQKQGGNN"
 12306..12944
 /gene="W50ERIPDM"
 /number=1
 complement(join(16037..16301,17591..17967))
 /gene="C75ERIPDM"
 complement(join(16037..16301,17591..17967))
 /gene="C75ERIPDM"
 /codon_start=1
 /product="PREG-like protein"
 /protein_id="CAC39033.1"
 /db_xref="GI:14140116"
 /translation="MAEGAEIYGESAHODMBRYAALAGILERYAGRNDAAAPPAEIA
 AAPAPRPRTAPGISVRYAARIARFAGSCSPACYVAYIYIDRLRRGRCLALALA
 VDSVYHRLITAVLSAVKPMDDICYNAAYPAKYGGISLAEMNYLEVDLFEGVFIDLN
 VYPERFAYCANLOSEMCAAPPRRLHCCSLDDGAGSSSSSLREAMEAS"
 complement(16037..16301)
 /gene="C75ERIPDM"
 /number=2
 complement(16302..17590)
 /gene="C75ERIPDM"
 /number=1
 complement(17591..17967)
 /gene="C75ERIPDM"
 /number=1
 complement(join(20929..21094,23560..23733,23808..23869,
 23964..24200))
 /gene="C90ERIPDM"
 complement(join(20929..21094,23560..23733,23808..23869,
 23964..24200))
 /gene="C90ERIPDM"
 /codon_start=1
 /product="WRKY-like DNA-binding protein"
 /protein_id="CAC39034.1"
 /db_xref="GI:14140117"
 /translation="MPLPPLDLPCHPKLLQMPFOEDOPGIHGVLSSDHGILPLPL
 ALPISAAAANAATVATIGKHSAAAGSMPIGAEVATVTYRAGNESITGSGTWMAG
 GSWAAGRGKMKIRKKRREPRFCFORSBYVDYDDGCKYKRYGQKYKKSILHRS
 FRTTHSNCVKKVERLSTDCRMVITTYEGRTHSPCDNDSGEHTNCFSSP"
 complement(20929..21094)
 /gene="C90ERIPDM"
 /number=4
 complement(21095..23559)
 /gene="C90ERIPDM"
 /number=3
 complement(23560..23733)
 /gene="C90ERIPDM"

intron
 /number=3
 complement(23734. .23807)
 /gene="C90BRIPDM"
 /number=2
 complement(23808. .23869)
 /gene="C90BRIPDM"
 /number=2
 complement(23870. .23963)
 /gene="C90BRIPDM"
 /number=1
 complement(23964. .24200)
 /gene="C90BRIPDM"
 /number=1
 complement(join(30589. .30712,31561. .31672,31787. .31850))
 /gene="C110BRIPDM"
 complement(join(30589. .30712,31561. .31672,31787. .31850))
 /gene="C110BRIPDM"
 /codon_start=1
 /product="putative protein"
 /protein_id="CNC39035.1"
 /db_xref="GI:14140118"
 /translation="MTTEVADQRMNAMDKVSAGCKGEVYVEEGQTFRRKKGIAGEL
 SSSPSIAPRRHAIHQRLGTGRSLDPVTPVLQARSAALSHWGEEVTSGL"
 complement(30589. .30712)
 /gene="C110BRIPDM"
 /number=2
 complement(30713. .31560)
 /gene="C110BRIPDM"
 /number=2
 complement(31561. .31672)
 /gene="C110BRIPDM"
 /number=2

Query Match	44.68;	Score 460.6;	DB 8;	Length 339972;
Best Local Similarity	78.18;	Pred. No. 1.4e-84;		
Matches 569;	Conservative	1;	Mismatches 150;	Indels 9; Gaps 1
QY 27	GGGGGTGCCACGGCAGCCGCTTGTAGCCCCCAATCCAGCTCCAAACCGCCGTTATCCG	86		
Db 193312	GGCCGGCCCGGTGACACACTTCTGGACACCGTGTCCCAAGCCCTCGAGAGCGTTCCGCGG	193313		
QY 87	GCAGTCCCAAGCTCACTGTCATCGCCTCCACTACTAATAATGCGCCCGGCTGGTCCAT	146		
Db 193312	CCCGAGCTGGGATCGCGGCTATATGCGCGGCCCGCTGTGAATGCCACACAGTGTCCCT	193253		
QY 147	C-----GTCCCCCGCGCGCTGTGCGCGGGGGGGGAGAGAGTCCGCGCGTCAAC	197		
Db 193252	GACCGCGCTGCGCCCGCGCGCTGCGCGGTGCGCGCTGGCGGGGAGCTCGACGTCCGCGG	193193		
QY 198	CGTTTCCACAGGAGTGTCTGTGTGTGGGCGACAAATATCGACACCGACAGATATCC	257		
Db 193192	CGTATTTCCACGGCCAGTGTCTTGTGTGTGGGGATTAATATGACACCGACAGATATCC	193133		
QY 258	CGCCGAGCACTCACTGTGTGTCCTCCAAACCGGAGAGATACCGCAAGTCCGTTCTT	317		
Db 193132	GCCCGAGCACTGACCTCGTGTCCCGTCCAAACCGGAGAGTACCGCAAGTCCGCTCGTT	193073		
QY 318	CGCTTCCGCGGGGCTCCCATCCGGGGGCTTACCGAGCCGTTGTGCTCCGGGTAGA	377		
Db 193072	CGCTTCTCTTGGCCCTCCCAACCGGGGCTTACCGAGCCGTTGTGCTCCCGCGGAGA	193013		
QY 378	GTCCTCCCGCTACGCATCATTTGTGCGGAGGACCACTTCGGGTGCGTTCTCTCGCA	437		
Db 193012	GACCAACCGGCTTACGCGGTGATCATTCGGGGGCGCAACTTTCGGGTGCGGTCTTCCGCGA	1929533		
QY 438	GCACGCGCCGCTGCGCTTGGGGCCGCTGGCGCACGCGCCATTTGTTCSSAAGGCTTACG	497		
Db 192952	GCACGCGCCGCTGCGCTTGGGGCCGCGCGCGCGCCCGCGCTGTGTCGCAAGGCTTACG	1928933		
QY 498	GGCATCTTTTTTGGCAACTCCGTGGCGACATGAGAGGTGATCCCTTGAGAGTCACGGA	557		
Db 192892	GGCATCTTTTTTGGCAACTCCGTGGCGACCGGAGAGGTGATCCCTTGAGAGTCACGGA	1928333		

ACCESSION AY035158 GI:14334969
 VERSION AY035158.1
 KEYWORDS FLI CDNA.
 SOURCE Arabidopsis thaliana.
 ORGANISM Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 1 (bases 1 to 970)
 REFERENCE 1
 AUTHORS Yamada,K., Liu,S.X., Sakano,H., Pham,P.K., Banh,J., Chung,M.K., Goldsmith,A.D., Lee,J.M., Quach,H.L., Toriumi,M., Yu,G., Bowser,L., Carninci,P., Chen,H., Cheuk,R., Hayashizaki,Y., Ishida,J., Jones,T., Kamiya,A., Karlin-Neumann,G., Kawai,J., Kim,C., Lam,B., Lin,J., Miranda,M., Narusaka,M., Nguyen,M., Palm,C.J., Sakurai,T., Satou,M., Seki,M., Shinn,P., Southwick,A., Shinozaki,K., Davis,R.W., Ecker,J.R. and Theologis,A.
 TITLE Arabidopsis Full Length cDNA Clones
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 970)
 AUTHORS Yamada,K., Liu,S.X., Sakano,H., Pham,P.K., Banh,J., Chung,M.K., Goldsmith,A.D., Lee,J.M., Quach,H.L., Toriumi,M., Yu,G., Bowser,L., Carninci,P., Chen,H., Cheuk,R., Hayashizaki,Y., Ishida,J., Jones,T., Kamiya,A., Karlin-Neumann,G., Kawai,J., Kim,C., Lam,B., Lin,J., Miranda,M., Narusaka,M., Nguyen,M., Palm,C.J., Sakurai,T., Satou,M., Seki,M., Shinn,P., Southwick,A., Shinozaki,K., Davis,R.W., Ecker,J.R. and Theologis,A.
 TITLE Direct Submission
 JOURNAL Submitted (10-MAY-2001) Plant Gene Expression Center, 800 Buchanan Street, Albany, CA 94710, USA
 COMMENT RIKEN Genomic Sciences Center (GSC) members carried out the collection and clustering of RAFL cDNAs (RAFL cDNA: 'RIKEN Arabidopsis Full-Length cDNA'): Seki,M., Narusaka,M., Ishida,J., Satou,M., Kamiya,A., Sakurai,T., Carninci,P., Kawai,J., Hayashizaki,Y. and Shinozaki,K.
 The Salk, Stanford, PGSC (SSP) Consortium members carried out the sequencing and annotation of the RAFL cDNAs: Yamada,K., Liu,S.X., Sakano,H., Pham,P.K., Banh,J., Chung,M.K., Goldsmith,A.D., Lee,J.M., Quach,H.L., Tang,C.C., Toriumi,M., Yu,G., Bowser,L., Lee,J.M., Quach,H.L., Tang,C.C., Toriumi,M., Yu,G., Bowser,L., Chen,H., Cheuk,R., Jones,T., Karlin-Neumann,G., Kim,C., Kosemra,E., Lam,B., Lin,J., Meyers,M.C., Miranda,M., Nguyen,M., Palm,C.J., Sakurai,T., Satou,M., Seki,M., Shinn,P., Southwick,A., Davis,R.W., Ecker,J.R. and Theologis,A.
 Yamada,K. (SSP/PGSC) and Seki,M. (RIKEN GSC) contributed equally to this work. Shinozaki,K. (RIKEN GSC) and Theologis,A. (SSP/PGSC) contributed equally to this work as PIs.
 FEATURES
 source location/Qualifiers
 1..970
 /organism="Arabidopsis thaliana"
 /db_xref="taxon:3702"
 /chromosome="2"
 /clone="RAFL04-10-K01 (R09710)"
 /note="This clone is in a modified pluescript vector (lambda ZAP) as a XhoI/SstI insert.
 ecotype: Columbia"
 1..970
 /gene="At2g43090"
 1..30
 /gene="At2g43090"
 31..786
 /gene="At2g43090"
 /codon_start=1
 /evidence-experimental
 /product="putative 3-isopropylmalate dehydratase, small subunit"
 /protein_id="AAK59662.1"
 /db_xref="GI:14334970"
 /translation="MAASIGANPTLSRTLASPKKPSFATFRSPPLRPNSTVASNF KPLVSREASSFVTRSAEPORBKTFHCIVKNDIDTDIOITPAEFLTPNSPEEY ERLGSLAVGLPASTKERFVPGEMKRTYIITIGENFGCCSSRHPAPYCLGAGGATA VVAOSYARIFRNVSATGEVPLDSEKAVCECTGDVATVELREGDSILNHTTGRE YLKLPIGDAGPVIDAGIIFAVARAKMIPSA"
 CDS
 5'UTR
 gene
 3'UTR

3'UTR 787..970
 misc_feature /gene="At2g43090"
 954 /gene="At2g43090"
 /note="artifact within poly A tail"
 BASE COUNT 269 a 210 c 211 g 280 t
 ORIGIN
 Query Match 20.4%; Score 210.8; DB 8; Length 970;
 Best Local Similarity 65.1%; Pred.No.2.2e-33;
 Matches 362; Conservative 0; Mismatches 182; Indels 12; Gaps 3;
 202 TTCACGCGAGTGTCTTGTGTGGCGACATATCGACACCGACCATATCCCGCC 261
 ||||| || ||||| || ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 238 TTCACATGACATGTGTATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 297
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 QY 262 GAGCAGCTGCTGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 321
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 298 GAGTTCTCAGCTGT 357
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 QY 322 TTCGCGGGGCTCCATCCGCGCTACCCGACGCGCTTCGTCGTCGTCGTCGTCG 381
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 358 TTAGTTGGCTTCCA---GCTTCTTACAAAGAAAGATTCGTTACAGCGATGAAG 414
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 QY 382 TCCCGCTACGCAATATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 441
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 415 ACGAAGTACTCATATCATATTCATGCGGTGAACCTTGTGATGTGATGATGATGAT 474
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 QY 442 GCGCGCGGCGGCTGTGGGCGCGCTGGCGACCGCATTTGCGSAGAGGTACGCGCG 501
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 475 GCTCCGCTTTGTTTGGACAGCGGAGCTAAAGCAGTGTGTGTGTGTGTGTGTGTGA 534
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 QY 502 ATCTTTTTCGCACTCCGTCGCGCACCTGTGAGAGGTATACCTTCGTGACGACGCTT 561
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 535 ATCTTTTTCGAGAACCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 591
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 QY 562 GGGGCTGTGAGAGGTGCAAGACAGAGGATGTGTGTGTGTGTGTGTGTGTGTGTGT 619
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 592 AGCGTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 651
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 QY 620 ----TTTATTTATACACACCTGTGTGACAGATGATGATGATGATGATGATGATGAT 675
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 652 AGTATTTGATCATATATATGACTGTGGAAGATGATGATGATGATGATGATGATGAT 711
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 QY 676 GGCCCTGTATATGAGCGGAGGATCTTCCCTACGCGGAGACAGATGATGATGATGAT 735
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 712 GACACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 771
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 QY 736 TCGAAGCTGTGAT 751
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 772 TCTGCTGCTGCTGAT 787
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 RESULT 6
 AC006224/c 45714 bp DNA linear PLN 27-FEB-2002
 LOCUS Arabidopsis thaliana chromosome 2 clone MFL8 map C1C10F02, complete
 DEFINITION sequence.
 ACCESSION AC006224
 VERSION AC006224.4 GI:20197626
 KEYWORDS HTG.
 SOURCE Arabidopsis thaliana.
 ORGANISM Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 1 (bases 1 to 45714)
 REFERENCE 1
 AUTHORS Lin,X., Kaul,S., Shea,T.P., Fujii,C.Y., Shen,M., Vanaken,S.E., Barnstead,M.E., Mason,T.M., Bowman,C.L., Rohning,C.M., Benito,M.-I., Carrera,A.J., Creasy,T.H., Buell,C.R., Town,C.D., Nierman,W.C., Fraser,C.M. and Venter,J.C.
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 45714)

AUTHORS Lin,X.
TITLE Direct Submission
JOURNAL Submitted (09-MAR-2000) The Institute for Genomic Research, 9712
Medical Center Dr., Rockville, MD 20850, USA
REFERENCE 3 (bases 1 to 45714)
AUTHORS Town,C.D. and Kaul,S.
TITLE Direct Submission
JOURNAL Submitted (27-FEB-2002) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA, cdrom@tigr.org
COMMENT On Apr 18, 2002 this sequence version replaced gi:6596536.
FEATURES
Source
1..45714
/organism="Arabidopsis thaliana"
/cultivar="Columbia"
/db_xref="taxon:3702"
/chromosome="2"
/map="C1C10P02"
/clone="MFL8"
complement(1..17781)
/note="overlap with BAC clone F14B2
(AC004450:1..53055).sequence between 11326 and 46597 in
F14B2 was deleted in MFL8"
complement(433..3077)
/gene="At2g43146"
/note="MFL8.1"
complement(join<433..671,762..1004,1193..1721,1803..2021,
2099..2166,2260..2325,2502..2583,2689..2741,2955..>3077))
/gene="At2g43146"
complement(join<433..671,762..1004,1193..1721,1803..2021,
2099..2166,2260..2325,2502..2583,2689..2741,2955..3051))
/gene="At2g43146"
/codon_start=1
/product="unknown protein"
/protein_id="AA15157.1"
/db_xref="GI:20197627"
/translation="MEALSLSTFGKSLPAIFAKKKLFFVYISGEDESDKLNRL
TWTDAVSADLSKCYILVHIOAGVDATNFSAIYPYSPICAIAGFSGQVWREGF
ITAEADLASLEKAMLGILQETIASIFSAIASQNETPSSASSVLPETGVPDAA
VASPTASVSQPESTSTVASTSTENNDGTAVAKESAEPSNLCDTTKNPASVD
GTKANVEHEATREPLRQAKKEPIPTAPGTNNTSVRSVDRKKQGTINEDSG
VGVSGRDINTLTKSVDFRTEKMKPKDEGEDEDEKSKASDVHLNTRLPGSSLOKPS
VTSLRMVKDYVSNQITIGADLAIVPPRYKTYTQDDLSLSELRFLDRALVYVP
RKRAITYQRTSYSESNNTDPNSGGTFATVVKVLSIANPFTFGGATNASSVPER
QTRPNTVRRNNLQVGTSTQDPSGSRNVNRRPTTSRLGSNIHTLNHEDAPPGDG
NAFWNGSTQYGGSGGSDNR"
complement(3457..5763)
/gene="At2g43143"
/note="MFL8.4: predicted by genscan and grail"
complement(join<3457..3658,3749..3826,3899..4276,
4359..4674,4768..4978,5095..>5763))
/gene="At2g43143"
complement(join<3457..3658,3749..3826,3899..4276,
4359..4674,4768..4978,5095..5763))
/gene="At2g43143"
/codon_start=1
/product="hypothetical protein"
/protein_id="AA15151.1"
/db_xref="GI:20197632"
/translation="MNPQOHLKLPKRLFFEPFLLSLYITTTIKTTISSQD
RHHPOLHAPSHSYSLPTSENRSPPLLPPSSSSSLSYFPCPNFNYL
PCHPSTARQYSIERHYRRERHCPDIAOEKFCIVKPYGYTPFPWPSRYAFNRN
VPFKRLAEKKTQWVRLLEGDRVFPFGGSPFGVGVADVILSVLPASGISITVL
DIGGVASGFLNLTNYKILFMSTAPRDIHQAQFALBERGLPAMGLVSTYELPRVL
SPDMVHSCRLVMTSYERFPYDGLIMEDVRLREBGVWLSGPVNASRYKFKNOK
RDSKELONOMEKLNDFRRRCFKRIASFPYVYWRKPSNHLQCRKLAKLPGLCSS
SDPDAATKMEPECLTFLPDVNDTNKTVLKNMERLNHPKMTGSIQCTTJAGFKAD
TNMQARVLYYDTRKFLSNGKRNVIDMNAQGGFAALIKYPMVMVMVAFEDKRN
TLGVVYDRGLIGTYMNCALSTYPTTYDI,IRANGVFSYLDKCDIVDILLMQRIILR
PEGAVIIRDFREDYVKKATITNOMRRNGTVPEDNSVFDHGTILIVDNSIK"
5518..5552
/rpt_family="(TGG)n"
complement(6380..6652)
/gene="At2g43136"

repeat_region
gene
/note="MFL8.5: predicted by grail"
complement(<6380..>6652)
/gene="At2g43136"
complement(6380..6652)
/gene="At2g43136"
/codon_start=1
/product="hypothetical protein"
/protein_id="AA15162.1"
/db_xref="GI:20197633"
/translation="MATTELKSAKYNKRRNRPSHGKRLKKKTYITTKKSKA
PTPDKTIEKSRSDQKTDNDEDSITHCWTRHGADKIRINMNNHR"
complement(7042..8055)
/gene="At2g43100"
/note="MFL8.14: supported by cDNA: gi_16974632_gb_AY060594
1."
complement(<7042..>8055)
/gene="At2g43100"
complement(7264..8034)
/gene="At2g43100"
/codon_start=1
/product="3-isopropylmalate dehydratase, small subunit"
/protein_id="AA15160.1"
/db_xref="GI:20197631"
/translation="MAVSPTRPQALPGSSTSSSLATFRSPFRNGSRIIPSS
SITSRGTSPTTIPRAASESDSNELATITFGICIVKDNITDQITIPAGACTFP
SNQDERDEIAHALSGLPDFHRTRFTEPGENSKYSIIIGENFGCSSRHAVACUG
AAQAKIVAESVRIFFRNSVATGVEFAPLESRCCECKGTGDTIELSDSGILLTN
HTTGKYNKLLKSIDAGPVIDAGGIFAYARMGMIPSLA"
complement(8391..9792)
/gene="At2g43090"
/note="MFL8.15: supported by full length cDNA:
Ceres:31507"
complement(join<8391..8537,8995..>9792))
/gene="At2g43090"
complement(9007..9762)
/gene="At2g43090"
/codon_start=1
/product="3-isopropylmalate dehydratase, small subunit"
/protein_id="AA15163.1"
/db_xref="GI:20197634"
/translation="MASSIOSANPLSLPLASPNKSSFAFRSPFLPENSYSANF
KPLVSRASSPTTIRSAAPROEKTRFGICVYGNITDQITIPREFITLVPNPEX
EKUGSTALVELPASTYKERVQPEMKTKYIIIGENFGCSSRHAVACILGAGAKA
VAVQSAARLEFRNSVATGVEYPLDSEVRVCDCTGDAVVELREGDSIILNHTTGKE
YKRLPIGDAGPVIDAGGIFAYARKGMIPSA"
complement(10120..12964)
/gene="At2g43080"
/note="MFL8.6: predicted by genscan and genefinder"
complement(join<10120..10227,10318..10410,10499..10588,
10837..10888,10982..11052,11145..11219,11452..11526,
11796..11861,12130..12182,12355..12411,12683..12730,
12901..>12964))
/gene="At2g43080"
complement(join<10120..10227,10318..10410,10499..10588,
10837..10888,10982..11052,11145..11219,11452..11526,
11796..11861,12130..12182,12355..12411,12683..12730,
12901..12964))
/gene="At2g43080"
/codon_start=1
/product="hypothetical protein"
/protein_id="AA15158.1"
/db_xref="GI:20197628"
/translation="MAPAMKYFGILTFYVGMVIGSILQALFIRNLRDSDYGTGPSL
RGIRGQVTRTRLRVSWADKDAELIRIGNVPEVYSRPRIYHDLSPCECEYLK
AIRRPLQSTVVDVKTGVSVDRTSSGMPLIVENSYPILQATIERIAVFSQVPA
ENDELQVLTLEPQDFYKPHHDYFATFNLKGGGVRVATMLVLDVDEGGETYPLPA
GDDCLQCGKIMKGISVKTPTKDAVLEWMSGLDGSDDPSRIGHGCEVLSCWASATKW
MRPKATS"
complement(13890..13916)
/rpt_family="(GAA)n"
13943..17829
/gene="At2g43105"
/note="MFL8.7"

```

/product="unknown protein"
/protein_id="AA01519.1"
/db_xref="GI:20197629"
/translacion="MSSFDPPNRYSAVLVILLIGESVAAADVSTEDSSLESPPC
TKRPMVKVILNWDVEGDFLTGLPAQGAALPSVPDALEPPAAFPVDPDSCSHSS
RUDGHIALSIRENCATFERAKAEAAAGASALLVINDKEDLMECMERKDIANSIYV
LMTSSGSDALAKMSVMDNKNVETLLAPKRPVADITAGILLIMAGTVVAANSSELT
DDQANESSTSLIAKDVSSNAGTRKDDPEEILDISYTGAVFFIYASIFELLPEFMS
NFWVWVLTFFICIGMGGMHNIIMAVILKRCILAKRSVLELGTMSVLSLIVLVC
AAVAAVEIKRHTSYSMVGODILGICIMTLQVNLPIKAVYLLCAAFYDITFWEF
ISPLFHEHSMVIAVQODSSTGESIPMLRIIPREFDPMVGDMIGFGDILPGLISIR
AARVSGSTLSSNPLRILAPLFLDRDKIKRVTISGVTFLMLTIGVIGLTLTYIGL
LAVILVLAVGELKEKELHNNGIESSHESTBEDPRPAA"
repeat_region
gene
complement(18357..18435)
/rpt_family="Ar_rich"
complement(18779..19472)
/gene="At2g43060"
/note="MFL8.8"
complement(<18779..>19472)
/gene="At2g43060"
complement(18853..18909)
/rpt_family="(78)"^n
complement(19002..19472)
/gene="At2g43060"
/codon_start=1

```

Best Local Similarity - 65.1%; Pred. No. 5e-33;
Matches 362; Conservative 0; Mismatches 182; Indels 12; Gaps 3;

9555 TTCCATGGACTGTCTATGTCTTCGGCGACACATCGACACTGACCAATCATTTCCGGCG 9496

9495 GAGTTTCTCACTCTCGTCCCTTCGAACTCAGAGGAATACGAGAACTCGGTCTTACGCT 9436

9435 TTAGTGGTCTTCCA---GCTTCTTACAAGGAACGATTGCTTCAGCCAGGTGAGATGAG 9379

9378 ACGAAGTACTCAATCATTCATTGGCGGTTGAAACTTTGGATGTGGATCGTCACGTCACAT 9319

9318 GCTCCGGTTTGTTAGGACAGCGGAGCTAAGACACTGGTGGCTCAGTCTTATGCTAGA 9259

Db 9258 ATCTTTTCAGCACTCTGTGCTACTGTGTAGGTTTATCCCTTGGATT---CTGAAGTT 9202

Db 9201 AGGGTTTGATGAGTTACAACTGGTGATGTTGCGACTGTTGAGTTGAGGGAAGCAGAT 9142

9141 AGTATTTCATCATCAGACTGGGAAAGAGTACAAAGCTTTAAGCCGATTTGGTATGCT 9082

RESULT 7				
AC004450				
LOCUS	90283 bp	DNA	linear	PLN 11-MAR-2002
DEFINITION	Arabidopsis thaliana chromosome 2 clone F14B2 map C1C10F02, complete sequence.			

ACCESSION	AC004450	GI:20197142
VERSION	AC004450.3	
KEYWORDS	HTG.	
SOURCE	Arabidopsis thaliana.	
ORGANISM	Arabidopsis thaliana	

REFERENCE	1 (pages 1 to 90283)
AUTHORS	Rounsley, S.D., Lin, X., Kaul, S., Shea, T.P., Fujii, C.Y., Mason, T.M., Shen, M., Rönning, C.M., Frisler, C.M., Somerville, C.R. and Venter, J.C.

REFERENCE 2 (bases 1 to 90283)

TITLE Direct Submission
JOURNAL Submitted (09-MAR-2000) The Institute for Genomic Research, 9712
Medical Center Dr., Rockville, MD 20850, USA

CONVNUM
JOURNAL
TITLE
DIRECT SUBMISSION
Submitted (27-FEB-2002) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA, cdtown@tigr.org
On Apr 18, 2002 this sequence version replaced qi:6598417.

Source	Location/Qualifiers
1	90283

misc_feature

gene

mRNA

CDS

TGCAT 751
 1 11
 ITGCAT 9006
 1 11
 3 GI:20197142
 90283 bp DNA linear PLN 11-MAR-2002
 is thaliana chromosome 2 clone F14B2 map C1C10F02,
 sequence.
 is thaliana
 Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Magnoliophyta; eudicotyledons; core eudicots;
 eurosids II; Brassicales; Brassicaceae; Arabidopsi
 1 to 90283)
 S.D., Lin, X., Kaul, S., Shee, T.P., Fujii, C.Y., Mason, T.M.,
 Rensing, C.M., Fraser, C.M., Somerville, C.R. and Venter, J.C.
 1 to 90283)
 Submission
 1 (27-FEB-2002) The Institute for Genomic Research, 9712
 Center Dr., Rockville, MD 20850, USA, ctdw@tigr.org
 , 2002 This sequence version replaced gi:6598411.
 Location/Qualifiers
 1. 90283
 /organism="Arabidopsis thaliana"
 /cultivar="Columbia"
 /db_xref="taxon:3702"
 /chromosome="2"
 /map="C1C10F02"
 /clone="F14B2"
 complement(1..53055)
 /note="Overlap with BAC clone MFL8
 (AC006224:1..17781), sequence between 11326 and 46597 in
 F14B2 was deleted in MFL8"
 complement(1..3839)
 /gene="At2g43105"
 /note="F14B2.1"
 complement(join(<1..405,565..585,777..919,995..1137,
 1476..1568,1641..1694,1898..2054,2140..2174,2276..2455,
 2671..2825,2947..3066,3157..3268,3398..3556,3690..>3839))
 /gene="At2g43105"
 complement(join(296..405,565..585,777..919,995..1137,
 1476..1568,1641..1694,1898..2054,2140..2174,2276..2455,
 2671..2825,2947..3066,3157..3268,3398..3556,3690..3839))
 /gene="At2g43105"
 /codon_start=1
 /product="unknown protein"
 /protein_id="AAI14939.1"
 /db_xref="GI:20197149"
 /translation="MSSFDPPHRRHSALVLLILLGFSVAADVDVSWEDSSLESPGC
 TNKRMVAVLIMVDSVEEDPGLTGAAGSALLVSPDOLRPAFVDEPLDSCSHLS
 RLDGSLASLINGSKVFTRAKIAEAGSALLVINDKEDIDMGCEKDTSLNVSIPV
 LMSGSDSLANKSVNCKNELLVYAPRPVVDLFAGILLIMANGTVVYASLMSLT
 PDDANESYSLAKDVSAGSKDPEKFIIDISVGAFFVTLSIFLLITFPPSS
 MEVWALTEFCFGQMGQHNIIIMAVILKRCVRLARKSVKPLDGTMSVLSLVNIVCL
 AFAVWFLIKRHSYVSGODIIGICIMLTALQVRLPNKIVATVLLSCAFVYDIFWV
 ISPLIFSHSNWIVVAGOSDGEISPMILIRIPFDPMGWIGIGDILFGILLISF
 ISRSFSTSLSNPLPRLLALPLFLDRDKIKKRVISNGYFILMTLTGYGIGILLTFYLG

repeat_region 3866..3892 LAVILGVLGELKELMWNVGEISESHTEPDEMPVA"
/rpt_family="(GAA)n"
gene 4818..7662 /gene="At2g43080"
/note="F14B2.2; predicted by genscan and genefinder"
mrna join(<4818..4881,5052..5099,5371..5427,5600..5652,
5921..5986,6256..6330,6563..6637,6730..6800,6894..6945,
7194..7283,7372..7464,7555..>7662)
/gene="At2g43080"
CDS join(4818..4881,5052..5099,5371..5427,5600..5652,
5921..5986,6256..6330,6563..6637,6730..6800,6894..6945,
7194..7283,7372..7464,7555..7662)
/gene="At2g43080"
/codon_start=1
/product="hypothetical protein"
/protein_id="AAC64297.1"
/db_xref="GI:3763917"
/translation="MAPAKTIVGELTFFVAVGAVTSGSLDIAFINLDELSTGTFPSL
KGLRGQNTRIKLDVSRMANKDDELKIGNVKEVVSPPRIIVLHDFLSPCECYLK
AIAARPLQVSTVVDVTKGVKSDVPTSGMFLIHLVRSYPIIQALEKRIAVSOVPA
ENGELIQVLRPEPOQFKPHHDYFADFTNLKRGQRYATMLATDVGGETPEPLA
GDGCTCGKIMKGISVYPTKGDVLFWSMGLDQSDPSRIHGCCEVLSGEKSATKM
MROKATS"
gene 7990..9391 /gene="At2g43090"
/note="F14B2.3; supported by full length cDNA:
Ceres:31507"
mrna join(7990..8787,9245..9391)
/gene="At2g43090"
CDS 8020..8775 /gene="At2g43090"
/codon_start=1
/product="3-isopropylmalate dehydratase, small subunit"
/protein_id="AAC64298.1"
/db_xref="GI:3763918"
/translation="MAASLOSANFTLSRTLASPKPSFATFRSPFLRENGTSVASNF
KPLVSREASSSEVTRSAEPOEKRTFGLIVVDNDIDDOIIPAEFLIVPEPSEY
EKLGSYALVGLPASYKERFVQPGEMTKYSIIIGENFGCGSSREHAPVCIAGAAGA
VVAOSYARIFRNSVATGEVPLDSEYRVCDCTGDVAVTELREGSITLINTTQKE
VKIRKIDGAFIVDAGGIFAVARAKMIPSAAS"
gene 9727..10740 /gene="At2g43100"
/note="F14B2.4; supported by cDNA: gi_16974632.gb-AV060594
1"
mrna 9727..10740 /gene="At2g43100"
CDS 9748..10518 /gene="At2g43100"
/codon_start=1
/product="3-isopropylmalate dehydratase, small subunit"
/protein_id="AAC64299.1"
/db_xref="GI:3763919"
/translation="MAASLPFFPOLPCSSNTKSSSLATFRSPFLRNNGSTLIPPSI
SITSGTSSPTIIPRAASESDSEALANTFGLICVLDNDIDDOIIPAGACAPF
SNOORDEIAHAHLSGLPDKFKRTIPRGNSKYSIIIGENFGCGSSREHAPVCLG
RAGAKATYAESYARIFRNSVATGEVPLSEVAVCECKTGDTITLSDSGLLTN
HTGKNYKLKSIDGAPVIDAGGIFAVARMGMIPSLA"
gene 12144..12220 /gene="At2g43137"
misc_RNA 12144..12220 /gene="At2g43137"
/product="Ath-chr2-sno27"
/note="snoRNA gene
http://rna.wustl.edu/snoRNAb/Arabidopsis/)"
gene 12301..12379 /gene="At2g43138"
misc_RNA 12301..12379 /gene="At2g43138"
/product="Ath-chr2-sno28"
/note="snoRNA gene
http://rna.wustl.edu/snoRNAb/Arabidopsis/)"
gene 12462..12540

misc_RNA /gene="At2g43139"
12462..12540 /gene="At2g43139"
/product="Ath-chr2-sno29"
/note="snoRNA gene
http://rna.wustl.edu/snoRNAb/Arabidopsis/)"
gene 12626..12700 /gene="At2g43141"
misc_RNA 12626..12700 /gene="At2g43141"
/product="Ath-chr2-sno28a"
/note="snoRNA gene
http://rna.wustl.edu/snoRNAb/Arabidopsis/)"
misc_feature 14493..15102 /note="molecular marker m551"
repeat_region complement(14897..14925)
/rpt_family="AT_rich"
repeat_region complement(15739..15760)
/rpt_family="AT_rich"
repeat_region 16763..16797 /rpt_family="AT_rich"
repeat_region complement(17020..17074)
/rpt_family="AT_rich"
repeat_region complement(18011..18045)
/rpt_family="AT_rich"
gene complement(18719..19970)
/gene="At2g43130"
/note="F14B2.7; identical to GB:D01026; supported by cDNA:
gi_217838.dbj.D01026.1,ATHARA4"
mrna complement(join(<18719..19272,19634..>19970))
CDS complement(join(18962..19272,19634..19967))
/gene="At2g43130"
/codon_start=1
/product="Ras-related GTP-binding protein (ARA-4)"
/protein_id="AAC64302.1"
/db_xref="GI:3763922"
/translation="MSDDDERGEYEFKVIILIGDSAVGRSNTLTRYARNEFNPNKAT
IGVFQNSMLIDGKEVKAQIQTAGGERRAVTSAYRGAVGALVYDITSSFFN
VGRMLDELNHSPTTVAKMLIGKKCDLESIRAVSVEGSLMESGLFPMETSLDST
NKRAPFEMVIREIYSNISKQLNSDSYKEELTVNRSYLVKNENEGKITPSCSR"
repeat_region 20107..20157 /rpt_family="AT_rich"
gene complement(21439..24702)
/gene="At2g43140"
/note="F14B2.8; predicted by genefinder"
mrna complement(join(<21439..21519,21599..21664,21808..21939,
22711..23012,24396..>24702))
/gene="At2g43140"
CDS complement(join(21439..21519,21599..21664,21808..21939,
22711..23012,24396..24702))
/gene="At2g43140"
/codon_start=1
/product="hypothetical protein"
/protein_id="AAC64303.1"
/db_xref="GI:3763923"
/translation="MYPNNSKSTAHDDGADDTNOYDSACATRDPSLGFQTHHP
PPQROQHOQNPMLVGHYLLGEPSSIGCFOSNASSSSSLFRHRSPPAGFDOLPDPN
GFSLGRPNGYGGGGEDGSPRLKSELFSSGSSSDHSHSLPISSEVAAAARNGVA
SSSMSPGNNTNMNDSSSHISFTIDQPGKRSKNSDFLETOYSWPOTTLMATMEN
LMNIPEDSVPCBARAKRGFAFTHPRSIARERRRIRSGKLKLLQELVPMNDKQTSADM
LDNIHEHKGIOHVEVRRP"
repeat_region complement(26301..26225)
/rpt_family="AT_rich"
repeat_region complement(29411..29451)
Query Match 20.4%; Score 210.8; DB 8; Length 90283;
Best Local Similarity 65.1%; Pred. No. 5.8e-33;
Matches 362; Conservative 0; Mismatches 182; Indels 12; Gaps 3;
QY 202 TTCACGGCGAGTGGTGTGGTGGGCGACAATATGCACACCGACGACATCATCCCGGC 261
DB 8227 TTCATGAGACTGTGCTATGCTGTCGGCGACAATCGACACTGCACCAATCATTCCTCCGC 8286

OY	262	GAGACACCACATCTGTCGGTCCCTCCAAAGCGGAGCAGTAACCAAGTCGTGCTTCCGC	321
Db	8287	GAGTTTCTCACTCTGTCGCCCTTCGAATCCAGAGAATACGAAACCTGGTCTTACCGCT	8346
OY	322	TTCGGGGGGCTCCCATCCGCGGGCTACCCGACGCCGTTTCGTGCTCGGGTGAGAGTCC	381
Db	8347	TTAGTTGGTCTTCCA---GCTTCTTACAGAGAACGATTCCTTACCGCAGGATGATGAG	8403
OY	382	TCCCGCTACGCCATCATTTGTCGGCGGAGCAACTTCGGGTGGGTTCCCTCCGCGAGAC	441
Db	8404	ACGAAGTCAATCAATCATATTGGCGGTGAATAAATTGGATGTGATGTCACTGTAACAT	8463
OY	442	GGCGCCGTGCGCTTGGGGCGCTGGCGCACGCGCCCATTTGTCGSGAGGGCTACGCGGC	501
Db	8464	GCTCGGTTTTTTTAGAGGACAGCGGGAGCTAAACACAGTGGTGGCTCATGCTTAATGTA	8523
OY	502	ATCTTTTTCGCACTCCGCTGGCCACTGGAGAGGGTGTACCCCTCGGAGCTACGAGAGT	561
Db	8534	ATCTTTTCAGAGAACTCTGTTGCTACTGCTGATGAGTTTATCTTTGGATT---CTGA	8580
OY	562	GGGGCTGGAAGGAGTGCAGACAAGGGGATGTGTCAACCGTGGACCTTGCTACTCCG--	619
Db	8581	AGGGTTTGTGATGATGTATACAACTGATGATGTGGCATGTTGATGAGGGAAGAAAT	8640
OY	620	-----TTTTTATTAAACCACTCTGTGCAAGGATACAAAGCTGMAACCAATTGGTANCT	675
Db	8641	AGTATTTTGATCAATCAATACATACGCTGGGAAAGATCAAAGCTTAAGCCGATTTGG	8700
OY	676	GGCCCTGTAAATGAGGCGGAGGAGATCTTTTCCCTACGCCCCGAGAGACAGATGATTCG	735
Db	8701	GGACCAAGTATGATGATGCTGGTGGTATTTGCTTATGCTTACGAAAAGCTGAATGATTC	8760
OY	736	TGCAAGCTGCTGCAT 751	
Db	8761	TCCTGCTGCTGCTGAT 8776	
RESULT 8			
Locus AY087084			
DEFINITION	AY087084	945 bp mRNA linear	PLN 26-JUN-2002
ACCESSION	AY087084	Arabidopsis thaliana clone 31507 mRNA, complete sequence.	
VERSION	AY087084.1	GI:21405808	
KEYWORDS	FIL_CDNA.		
SOURCE	thale cress.		
ORGANISM	Arabidopsis thaliana		
REFERENCE	Haas,B.J., Volfovsky,N., Town,C.D., Troupkan,M., Alexandrov,N., Feldmann,K.A., Flavell,R.B., White,O. and Salzberg,S.L.		
AUTHORS	Full-length messenger RNA sequences greatly improve genome annotation		
TITLE	Genome Biol. (2002) In press		
JOURNAL REFERENCE	2 (bases 1 to 945)		
AUTHORS	Broyer,V., Troupkan,M., Alexandrov,N., Lu,Y.-P., Flavell,R. and Feldmann,K.		
TITLE	Full-length cDNA from Arabidopsis thaliana		
JOURNAL REFERENCE	Unpublished		
AUTHORS	3 (bases 1 to 945)		
TITLE	Direct Submission		
JOURNAL REFERENCE	Submitted (11-MAR-2002) Ceres, Inc, 3007 Malibu Canyon Road,		
AUTHORS	Malibu, CA 90265, USA		
COMMENT	This clone sequence is one of 5,000 Ceres full-length cDNAs made available to TIGR and Genbank. The following quality assessment of this set was done by comparison with known proteins: two percent of the clones are estimated to be 5'-truncated; less than one percent are 3'-truncated; approximately two percent represent alternative splice variants, including unspliced introns and spliced exons; one		

percent may contain premature stop codons: five percent may have frame shifts in a coding region. A sequence is considered to be 5'-truncated if it lacks the translation initiation start (ATG). A sequence is considered to be 3'-truncated if it lacks the C-terminal end of the encoded protein. Please note that these cDNA sequences are derived from the Ws or Laver ecotypes and therefore may contain polymorphisms when compared to sequences from Col-0. Genest carried out the library production and sequencing of the full-length clones. Ceres, Inc. carried out the clustering of the 5 sequences, selection of clones, and sequence assembly.

FEATURES
source

Location/Qualifiers
1. .945

CD

CC
IN

0374

57

2

2

2

2

3

3

3

1

•

1

3

U

5

תא

ת

①

2

5

1

7

7

FEATURES		source	
<p>percent may contain premature stop codons; five percent may have frame shifts in a coding region. A sequence is considered to be 5'-truncated if it lacks the translation initiation start (ATG). A sequence is considered to be 3'-truncated if it lacks the C-terminal end of the encoded protein. Please note that these cDNA sequences are derived from the Ws or Iwer ecotypes and therefore may contain polymorphisms when compared to sequences from Col-0. Genset carried out the library production and sequencing of the full-length clones. Ceres, Inc. carried out the clustering of the 5' sequences, selection of clones, and sequence assembly.</p>		<p>Location/Qualifiers</p>	
<p>1. .945</p> <p>/organism="Arabidopsis thaliana"</p> <p>/db_xref="taxon:3702"</p> <p>/clone="31507"</p> <p>31. .786</p>		<p>/codon_start=1</p> <p>/product="3-isopropylmalate dehydratase, small subunit"</p> <p>/protein_id="AAM64645.1"</p> <p>/db_xref="GI:21592696"</p> <p>/translation="MAASLOSANPTLSRTLASPNKPSFATFRSPLEFNSTVASNF KPLVSRASSFLVRSAPDPERKFTGLVVDNDIDPOIIIPAEFLTVSPDEY EKIGSYALVGLIPASYSKERFVPOGEMKTKYLIIGENRGGSSREHAPVGLAAGAKA VVQSRIRIFRNSVATGEVYPLPDESRVDECTGVAIVELREGSILNHTTKE YKKPIGDAGPVIDAGGIFAFIRARAGITPSAA"</p>	
<p>BASE COUNT 246 a 208 c 210 g 281 t</p> <p>ORIGIN</p>		<p>20.3%: Score 209.2; DB 8; Length 945;</p> <p>Best Local Similarity 64.9%: Pred. No. 4.8e-33;</p> <p>Matches 361; Conservative 0; Mismatches 183; Indels 12; Gaps 3;</p>	
QY	202 TTCCAGCGCGAGTGGTGGTGGCGGCGACATATCGACACCGACAGATCATCCCGCC	261	
Db	238 TTCCATGACGATGTGCTATGTCTGGCGGACCAACATCGACATGACCAATATTCGCCG	297	
QY	262 GAGCACCTCACTCTGTCGCCCTCCCAAGCGGCGAGTGATCCGGAAGCTCGGGTTCCTCGCC	321	
Db	298 GAGTTCTCACTCTGTCTCCCTTCGATCCAGAGGAATACGAGAACTCGGTTCTTACCT	357	
QY	332 TTCCGCGGGGCTCCCATCCGCGGCGCTACCCGAGCGCGTTGCTGCTCGCGGTGAGAGTCC	381	
Db	358 TTAGTTGGCTTCCA--GCTTCTTAAGAGAACGATTCGTTCAAGCGAGTGAGATGAG	414	
QY	382 TCCCGCTACGCCATATGTGGGCGGACCAACTCGGGTGGGTTCCCTCGCGAGGC	441	
Db	415 ACGAAGTACTCATATATATGTGGCGGTAAAACTTTGGATGTGGATTCGTACAGTAAAT	474	
QY	442 GCGCCCGTGGCGCTTGGGGCGCGTGGCGACGCGGCATTTGTCGAGAGGCGCTACGCGCGC	501	
Db	475 GCTCGGTTTGTATTAGGAGCAGCGGAGACTAAAGCAGTGGTGCGTCACTCTATATCTTGA	534	
QY	502 ATCTTTTTCGCACTCCGTGGCCACTGAGAGGTGTACCCCTCGAGACTACGGAGGTT	561	
Db	535 ATCTTTTCAGAACTGTGTGTACTAGTGGTGAAGTTTATCTTTGGATT--CTAAGATT	591	
QY	562 GGGGCGTGAAGAGATGCAACAGAGGAGTGGTGCACCGCTGGACCTGTACTACCG--	619	
Db	592 AGGTTTGTGATGATGTACAACTGTGTATGTGGAGCTGTGAATGAGGAAGAGAT	651	
QY	620 ---TTTATTATACCAACACTCTGGCAAGAGTTCAAAGCTGAACCAATTGGTATGCT	675	
Db	652 AGTATTTTGTATCAATCATACGACTCGGAAAGGTATTAAAGCTTAACCGATTGGTATGCT	711	
QY	676 GGGCGCTGAATGAGGCGGAGGAGTCTTGGCTATCGCGCGGGAAGACAGAAATGATTCG	735	
Db	712 GGACCAAGTATGATGCTGGTGGTATATTTTGCTTATGCTAGGAAGCTGGATGATTC	771	
QY	736 TCGAAGCTGCTGCAT 751		
Db	772 TCTGCTGGCTGTGAT 787		

Arabidopsis cDNA clones
Unpublished
2 (bases 1 to 1014)
Chen, R., Chen, H., Kim, C.-J., Koesema, E., Meyers, M. C., Ban, J.,
Bomser, L., Carninci, P., Dale, J. M., Goldsmith, A. D., Hayashizaki, Y.,
Iwata, J., Jiang, P. X., Jones, T., Kamli, Y., Karlin-Neumann, G.,
Kawai, J., Lam, B., Lee, J. M., Lin, D., Liu, S. X., Miranda, M.,

intron complement(749..826)
/gene="F17J16_10"
/number=1
exon complement(827..967)
/gene="F17J16_10"
/number=2
intron complement(968..1123)
/gene="F17J16_10"
/number=2
exon complement(1124..2038)
/gene="F17J16_10"
/number=3
gene 2890..4538
/gene="F17J16_20"
complement(join(2890..3456,3540..3955,4034..4336,
4433..4538))
/gene="F17J16_20"
complement(join(2890..3456,3540..3955,4034..4336,
4433..4538))
/gene="F17J16_20"
complement(join(2890..3456,3540..3955,4034..4336,
4433..4538))
/gene="F17J16_20"
/note="similarity to mitochondrial RNA splicing protein
MRS2 - Saccharomyces cerevisiae, PIR:S62064"
/codon_start=1
/product="putative protein"
/protein_id="CAB86925.1"
/db_xref="GI:7529740"
/translation="MAEGFTIGIGSEFEAEATGAFSYSHHQSIDLRSSAVKDSA
QTSTSPSPPLPIHAGSAGATGAKKKGARLWMDRDCAMEVVCDSSTIK
RASVPARDRLIGPVFSSHSNIGKNEDLLPGIARAKAIVNLEVIKAVLEAYLL
LDPRLPEVLPPEERLKQOPQNGNENALQASAVOSLPDEAEAGSLSLPEPFVYL
ETALEVCSFVDSVALETAMPIVDELTKNVTENLEYSRSLKSNLTSLRLVQVK
RDELEHLDDNEDMADLYLTKRNIONQTEALAGTASNSIALPAHNTSLNHLTSR
SASMTNSTNEEDVDLEMLLEAVFMQDGRKMLIVREVYIDTPEYVNICDQNR
ETIOQLTITASEPAIAETLLASFGNITPCPLSIHGVPGRVMSVTALCTVLEAV
TIGIARMKKLGSS
exon complement(2890..3456)
/gene="F17J16_20"
/number=1
intron complement(3457..3539)
/gene="F17J16_20"
/number=1
exon complement(3540..3955)
/gene="F17J16_20"
/number=2
intron complement(3956..4033)
/gene="F17J16_20"
/number=2
exon complement(4034..4336)
/gene="F17J16_20"
/number=3
intron complement(4337..4432)
/gene="F17J16_20"
/number=3
exon complement(4433..4538)
/gene="F17J16_20"
/number=4
gene 5601..10237
/gene="F17J16_30"
complement(join(5601..5954,6085..6186,6264..6481,
6560..6832,6919..7108,7227..7242,7680..8140,8590..8850,
8953..9102,9329..10237))
/gene="F17J16_30"
complement(join(5601..5954,6085..6186,6264..6481,
6560..6832,6919..7108,7227..7242,7680..8140,8590..8850,
8953..9102,9329..10237))
/gene="F17J16_30"
/note="similarity to several hypothetical proteins -
Arabidopsis thaliana
contains EGF-like domain signatures AA367-379;Prokaryotic
membrane lipoprotein lipid attachment site AA369-379"
/codon_start=1
/product="putative protein"
/protein_id="CAB86926.1"

/db_xref="GI:7529741"
/translation="MDRIINLPEIICHIVSFLSAKEAPASILSKMRNLFTVIKIL
QFDDSVKNESLKDQVDGVALPTSSVRSCSLECRREDPPHYDFNFCICALKRG
ILDLKLDICAGRRYSLEFVETCKTVLKEIGSDGFVVDVPEDAFIPALETLIN
YIRFKDLRCAAFERKLSACVLKEIYHNMEMRMRSGNISPTLERLTISHVDLYE
CEFTIRINDPPNLTLYELSDAVPDDPIYVNLDSIAEVKLDITLMDVKHGVVDNDT
ISSNPTNINIGLRVETIMNOSPNTQAPSPYFEALPVKRIKXLTITRNDDVFCG
EFLPVIKCCPNLKTLYIDPPLHYNEDRPKSVCHGSSGSLISCLPLEVLTQITDSSGT
FGEVEQLKHFLEKLSGLKVLKSLTRFGSDKRLMLPRASSCKTKHYDSLENAL
PSLKTLLIDSVKFEYDRGCCAFQKLISACVLVESVVRNLEMDWESGSCASSQTLER
LTIDHRYWAENHLESFTFDPSSLTYLDYNAHVGSPTVNLDSLEAKLNGFTTDLV
EDDDPPTSPDPTNLKGLRVNELLRLMIVVGNFESRVOKVDLSNLPPEYCHITSE
LPTKHAALTSVLSKSNLMLKVLPNLIDIDSESLPEERLPEELSRITVEKIKSER
YVDMRRDRGVLPMLKTLIDIDSESHCGTIDMFLPSFVLEILKRVANEMWESDETV
VSATIKLISHTGVAVEDYPLVNNKALFEARINFTVEKQIKRVRAENRMLQDDDEV
VIRFSNVYKLMANGONOKXLTADTADPEARINFTVEKQIKRVRAENRMLQDDDEV
MPVGLHMTDKGDCDCISREDGRSLTSCPVKLEIRGRGTRKMTKIKHFLV
YFCLKEKNIVVEYDPTELRFQVSKVLIQMEEVNKLSCNVOLLITDYLSKMTA
AGRL"
exon complement(5601..5954)
/gene="F17J16_30"
/number=1
intron complement(5955..6084)
/gene="F17J16_30"
/number=1
exon complement(6085..6186)
/gene="F17J16_30"
/number=2
intron complement(6187..6263)
/gene="F17J16_30"
/number=2
exon complement(6264..6481)
/gene="F17J16_30"
/number=3
intron complement(6482..6559)
/gene="F17J16_30"
/number=3
exon complement(6560..6832)
/gene="F17J16_30"
/number=4
intron complement(6833..6918)
/gene="F17J16_30"
/number=4
exon complement(6919..7108)
/gene="F17J16_30"
/number=5
intron complement(7109..7226)
/gene="F17J16_30"
/number=5
exon complement(7227..7242)
/gene="F17J16_30"
/number=6
intron complement(7243..7679)
/gene="F17J16_30"
/number=6
exon complement(7680..8140)
/gene="F17J16_30"
/number=7
intron complement(8141..8589)
/gene="F17J16_30"
/number=7
exon complement(8590..8850)
/gene="F17J16_30"
/number=8
intron complement(8851..8952)
/gene="F17J16_30"
/number=8
exon complement(8953..9102)
/gene="F17J16_30"

Query Match 19.0%; Score 196.6; DB 8; Length 89517;
Best Local Similarity 60.3%; Pred. No. 5e-30;
Matches 382; Conservative 1; Mismatches 235; Indels 15; Gaps 3;

QY	114	CCACTCCTAAATGCGCGCGCTGGGTGATGTCCTCCCGCGCGCTGCTCCCGCGC	173
DB	11388	CAAAACCACTTACTTCTCTCCCGCAGATCATACAGGGGTGCTGCATCATCCCGCA	11329
QY	174	GGGACGACAGCTGCGCGCTGTCAGCGCTTTTTCACGAGAGTGTGGTGGAGCA	233
DB	11328	TTTCAGCGCGATCAATTAACAGAGACTTTCACAGGCGCTGCTGCTGTAAAGACA	11269
QY	234	TATCGACACCGACAGATATCCCGCGGAGACACTACTCTGTGCTCCCAAGCCGA	293
DB	11268	CATCGACCGATCAAAATAATCCCGCGGATGAGGAGTCTCATCTCGATTCAGA	11209
QY	294	CGAGTACCGACAGCTGCGCTTTCGCTTCCGCGGCTCCCATCCGCGCGACCCGAC	353
DB	11208	AGATGCGGAGAACTCGGCTCTTTCGCGCTTAACGCGTTACCA--AAATCTACACGA	11152
QY	354	GCGCTTCTGCTCGCTGAGAGTCTCCCGCTACGCGCATATGTCGGGAGCCAA	413
DB	11151	ACGTTTCGTTTCCAGAGAGATGAATCAAACTCATCATCATCGCGCGCATTA	11092
QY	414	CTTCGGTCCGCTTCTCTCGCGAGACGCGCGCTTGGGCGCGCTGCGCGACG	473
DB	11091	TTTCGGTTCGGATCTTCCCGGAGAACGCTTCAGTTTGTCTGCGCGCGCGGAGCTTA	11032
QY	474	CGCATTTGTCGAGAGGCTACGCGCGATCTTTTTCGCACTCCGTCGACCTGAGGA	533
DB	11031	ACCTGTGTGGGAGATGCTAGCTAGATCTTTTTCAGAGATGCTAGCTAGAGTGA	10972
QY	534	GGTGTACCTCTGAGACTCAGGAGCTGTGGGCGCTGAGAGAGTGCAGACAGGGATG	593
DB	10971	GATTTTCCGTTGGA--ATCGAGAGTTAGATTTGGAGAGTGCAGAAACAGGGATGT	10915
QY	594	GCTCAC-----CGTGACCTTGCTACTCGCTTTTATTAACCAACCTCTGCGCA	644
DB	10914	GGTGACATCGAACCAAGAGAGAGAGATGATGCTGAGCTGATTAATGAGGAGGATCTT	10855
QY	645	GGAGTACAGCTGAACCAATTTGATGCTGAGCTGATTAATGAGGAGGATCTT	704
DB	10854	AGATATCAAACTGAACCGCTCGGAGTGGCGGTGATGATGATGATGATGATGATGAT	10795
QY	705	TGCGTACGCGCGAGAGAGATGATGCTC	737
DB	10794	CGCTTATGCAAGAAAGCGCGATGATCTTC	10762
RESULT 14			
AE012835/c			
LOCUS	AE012835	10029 bp	DNA linear BCT 01-JUL-2002
DEFINITION	Chlorobium tepidum TLS	section 56 of 194 of the complete genome.	
ACCESSION	AE012835		
VERSION	AE012835.1	GI:21646549	
KEYWORDS			
SOURCE	Chlorobium tepidum TLS		
ORGANISM	Bacteria; Chlorobi; Chlorobia; Chlorobiales; Chlorobiaceae;		
REFERENCE			
AUTHORS	1 (bases 1 to 10029)		
	Eisen, J.A., Nelson, K.E., Paulsen, I.T., Heidelberg, J.F., Wu, M.,		
	Dodson, R.J., Deboy, R., Gwinn, M.L., Nelson, W.C., Haft, D.H.,		
	Hickey, E.K., Peterson, J.D., Durkin, A.S., Kolonay, J.L., Yang, F.,		
	Holt, I., Umayam, L.A., Mason, T., Brenner, M., Shea, T.P., Parksey, D.,		
	Nierman, W.C., Feldblyum, T.V., Hansen, C.L., Craven, M.B., Radune, D.,		
	Vamathevan, J., Khouli, H., White, O., Gruber, T.M., Ketchum, K.A.,		
	Venter, J.C., Tettelin, H., Bryant, D.A. and Fraser, C.M.		
TITLE	The complete genome sequence of Chlorobium tepidum TLS, a		
JOURNAL	photosynthetic, anaerobic, green-sulfur bacterium		
PUBMED	Proc. Natl. Acad. Sci. U.S.A. 99 (14), 9509-9514 (2002)		
AUTHORS	2 (bases 1 to 10029)		
	Eisen, J.A., Nelson, K.E., Paulsen, I.T., Heidelberg, J.F., Wu, M.,		
	Dodson, R.J., Deboy, R., Gwinn, M.L., Nelson, W.C., Haft, D.H.,		
	Hickey, E.K., Peterson, J.D., Durkin, A.S., Kolonay, J.L., Yang, F.,		

TITLE	JOURNAL	REMARK
Chlorobium tepidum TLS	Submitted (30-APR-2002) The Institute for Genomic Research, 9712	
Direct Submission	Medical Center Dr, Rockville, MD 20850, USA	
unpublished		
FEATURES	source	location/qualifiers
gene	1..10029	/organism="Chlorobium tepidum TLS"
	/strain="TLS"	
	/db_xref="taxon:194439"	
	complement(90..1694)	
	/gene="TleuA-1"	
	/note="CT0612"	
CDS	complement(90..1694)	
	/gene="TleuA-1"	
	/note="Identified by match to PFAM protein family HMM	
	PF00682"	
	/codon_start=1	
	/transl_table=1	
	/product="2-isopropylmalate synthase"	
	/protein_id="AAW1854.1"	
	/db_xref="GI:21646550"	
	/translation="WVLELYDTLRDGTGEGHINLSVODKLLIARLDEFEVDIEGG	
	WPSSNPFDEEFELKARKLNLKARLALFASTASLDNENDNGLVRCAPVLTLE	
	GWTAKHVSYSLSIGSDENAEILYRSVKELTVSGRVEFDADHFDGKMDAGAEEM	
	IAAIVDGGASRVVLCIDNGCTLPHEIAALIPVRELIGVSGIHAHNSDILAVNSIE	
	AVRAGAVOQGTINGIGERCANLIVSIIPNMLKIGAFESHVDLKEITVSKFVE	
	ILNLPDPSKAPFEKSAFAKGGIHYSAVKKESSLYEHDPLVGNRVRVSLAG	
	SNTRYKQDELGISLPEKGEVFKLVVHVKIKIHOGIOPGDAASFELLRLRELQFP	
	YEVLESVYVIONCOELKAYDAQVMVMGDEPQVADGCPVNAIDAKRAKLLIF	
	LPDIRMTLIDYKRVLEKSGTSARVFLIESDSQNSMGTVGVSTITIASQLALN	
	DSINLYLFYMQSKAAATAPASBALNS"	
gene	complement(1703..2275)	
	/gene="CT0613"	
CDS	complement(1703..2275)	
	/gene="CT0613"	
	/note="Identified by match to PFAM protein family HMM	
	PF00694"	
	/codon_start=1	
	/transl_table=1	
	/product="3-isopropylmalate dehydratase, small subunit,	
	putative"	
	/protein_id="AAW1855.1"	
	/db_xref="GI:21646551"	
	/translation="MDRIIGKAYVYLGNKINIDDOIIPAEHLVYSLSDEPEYKMGKXA	
	LSGVPIQAGLPREGNIPVVEGEFTSPSIIITAGPFGGSGRREHAPALVAGAKAI	
	IAESYAHIFRNCYDGSFVIFPETAQPLKKSIMTGDLSLDMENNTLNLQNTIYEL	
	RPIDGVINIVDGGIIFETARKNMLASTEA"	
	complement(2311..3606)	
	/gene="CT0614"	
	complement(2311..3606)	
	/gene="CT0614"	
	/note="Identified by match to PFAM protein family HMM	
	PF00330"	
	/codon_start=1	
	/transl_table=1	
	/product="3-isopropylmalate dehydratase, large subunit,	
	putative"	
	/protein_id="AAW1856.1"	
	/db_xref="GI:21646552"	
	/translation="MAOTIIOKIFBARANKRFVDPGQSVMLNVDLTLHDVCGPFTD	
	IFKQEPGNKAWPDSKVVYLPRPHYIFANEHAHNIDILROFAAEOGLPAYVYVGD	
	IRGCVHVALAEGSPNPGVILPQDSTHSCGAGPFGGSGRREHAPALVAGAKAI	
	KVPSKMTFREGQPEPLRKADILIIDITDITDQATRAMPEQEAAYSIPIDERMF	
	LCNNAIEAGGANGITADAVTEAFVAKRSKPELFTSPDPAQIHSMTVAATNIVA	
	VAKHSPDNRAIVHSVAGPEITKSYIGCTGKLTDFPKLAALIKGKRVAAVTNIVA	
	TLVAASQLETEMVDGQTLRIIEFACNIALPSCACACLGSPDITVGRSVDMVVVST	
	NRNPPGMSKFAVSYLASFPLTAAASATIGKLTDPDFL"	
	complement(3775..4833)	
	/gene="TleuB"	

DEFINITION Alcaligenes eutrophus 4-hydroxybutyrate dehydrogenase (ybd), ORF 2 and 4-10 genes, complete cds, and ORF3 and 11, 3' end.

ACCESSION L36817

VERSION 1.0

KEYWORDS 4-hydroxybutyrate dehydrogenase; NAD-dependent; lipase.

SOURCE Alcaligenes eutrophus (strain H16 SK4040) DNA.

ORGANISM Ralstonia eutropha

Bacteria; Proteobacteria; beta subdivision; Ralstonia group; Ralstonia.

REFERENCE 1 (bases 1 to 9830)
Valentin, H.E., Zwillingmann, G., Schonebaum, A. and Steinbuechel, A.
Metabolic pathway for biosynthesis of
poly(3-hydroxybutyrate-co-4-hydroxybutyrate) from 4-hydroxybutyrate
by Alcaligenes eutrophus

JOURNAL Eur. J. Biochem. 227 (1-2), 43-60 (1995)

MEDLINE 95154322

PUBMED 7851418

FEATURES

source

1. .9830 Location/Qualifiers

/organism="Ralstonia eutropha"

/strain="H16 SK4040"

/db_xref="taxon:510"

<1. .164

/gene="ORF3"

/codon_start=2

/transl_table=11

/protein_id="AAC41419.1"

/db_xref="GI:703067"

/translation="NSVGNLAGFVSPYVWALKDMTOSTQLPMYVLSALIVGAVLVW
LTPAKLVNR"

1. .163

/gene="ORF3"

1/0. .204

/note="putative"

222. .225

232. .1014

/note="ORF4"

/codon_start=1

/transl_table=11

/protein_id="AAC41420.1"

/db_xref="GI:695274"

/translation="MPRFANLSMNYNEHAFIDRFAAAAADGFAVEELFPYEHAAE
LRALDANGLTQALFNAAPGDMAAGERGLAALFGRADFRGTIGRALFAGVIGNDRI
HWAGLIPADPARRCARATYLENLAFAANAAAGVTLEIPINTRDMGYLENRDD
GOATCEVGGANLKVQDCYHCOIVEGDVMAKLRDIAIGHIQLAGVDEPHEPDVGE
LNPYLFVYVDITGIDGMIGCEYRPAAGTSAGLGMKPYLGR"

1053. .2027

/note="ORF5"

/codon_start=1

/transl_table=11

/protein_id="AAC41421.1"

/db_xref="GI:695275"

/translation="MNYLTGGAGFLGIOLARLLRGTLNLDGQVATKRLTLDVY
APGOLDARVRYTGDISDPAYLRQALIDTGTGAVPHLAANYSCQARADDDIGRVNLD
ASRALITGCELGHPRYLFTSSVAYGGQLPVPVODTALNQSSTGCVOKAIGELL
SDYSRRGFVDGRVLRLEPTISVRGKPMNAASPSASGIIIRPLSGVANCVAPETPLM
LISRAVAALVINGIELAGERLGNRRVYNLPGLSVPAAGMTELRRAVAGNAVADRVTW
EREARVENIVGTMPAAWNAERALLAFQSDASDEVIRAYMEDAGLAK"

complement(2076. .2107)

/note="putative"

complement(2131. .2155)

/note="putative"

complement(2143. .2580)

/note="ORF6"

/codon_start=1

/transl_table=11

/protein_id="AAC41422.1"

/db_xref="GI:695276"

/translation="MAKEDFRHSITPLRVAEVDPOSIVENAHLYLYCICVETVRA
LGIRPEVDVLAHGVDFIVVKSLEYHASAFDDELIRGMSRLGRSSMLFRVEMER
GDEHLITGEIYVVCADPATOKSAIIPGVREIIIEGYEVVPAG"

2590. .2595

RBS complement(2619. .2665)

terminator

terminator

CDS

/note="putative"

complement(2675. .2722)

/note="putative"

complement(2692. .2706)

/note="-24/-12 signal"

complement(2697. .3680)

/note="ORF7"

/codon_start=1

/transl_table=11

/protein_id="AAC41423.1"

/db_xref="GI:695277"

/translation="MQRIHFARAGIAAATAALGIAAMPQAQADKFPQRPRLVIGY
TAGSSTDIIPRVLADNAKILIGQPIVENKPGAGVLPQPMQSTAPDQYTAQVAMP
VYRLPYTKITNMPVDKLYIINLAGSEGLVVPADSPKTMQETAYVAKKANGKLT
GSPKSMTLTKITMBELAMKQGVFQSHPKRGNSEMOALLGHHVMSVADTPAMAPVE
QGKRLISTWGEKRSARFSPVPTIKELGICITQTSPPGIVAPKGTQPKIVOKLHDAFK
KAMDPNYRESLAKFDMEPYTKNSQOYQAPAAETVKKKAIIEKGLAAQ"

complement(3691. .3696)

complement(3762. .4856)

/note="ORF8"

/codon_start=1

/transl_table=11

/product="lipase-like enzyme"

/protein_id="AAC41424.1"

/db_xref="GI:695278"

/translation="MSRCGRRAELHNAAGTSGSASFPMKSAFSSPDAPAPFGLSS
PLDPQVALLLELIRAKRPPILHAFEPDAKIAVEKSNPILIDIPPPYMAEDLLAPAR
DGHAIPRLRYGSAWTEPILPVYHGGGTFVGSVSDHPLCLRLCGQADCMVASR
DRIAGPQMPREPPTANDAPVILHFAFGRLGADPARIAVGSAGSTGLAAACAVAR
NAGLAPVQLITITGTCARQDTPSHRLAGCYLLTMDIMRFPQAYLDOEASDDMR
APLDGGAGAEVGTCPAMIAVAGTDFLHDBGYAIAEKLAAGVAATLADYPMIHDF
FKLGRFPVAVAAQAAHAEVAALRAAFGTPHN"

4853. .4859

/note="-43 region"

4864. .4869

4888. .4893

complement(4941. .4955)

/note="-24/-12 signal"

5105. .5109

5120. .6268

/gene="ybd"

5120. .6268

/gene="ybd"

/EC_number="1.1.1.61"

/codon_start=1

/transl_table=11

/product="NAD-dependent 4-hydroxybutyrate dehydrogenase"

/protein_id="AAC41425.1"

/db_xref="GI:695279"

/translation="MARIYVLTHTHIDGAVSLKSECEIRIGRRPLLYTDKGVVAG
VAQRIDAMQQLQAVNPDETSNPTPEANVRRAAOYREAGDCGLVAVGGSSIDLAKG
IALATHEGELTYATTIEGSSARITDKAAPLIAYPTTSGTGESPARCAIITLIDGRKL
GHSWHLPKSAVCDPELITGLPGLTATATGMDIARLIEFFLPAFNPADGIALDG
LERGGHIERATRDQDARDARLNMSASMOGAMFOKGLGCVHSLSPGLKIDGRT
GLHGTILNAVMPAVLRNNADAPTVYRDDRYARLRRAHNLFDGADIAQVADMTVRUG
LPTGRQMGVTEMDKVTIAGLVLDHCKRTKPKBASADYRRMLEQSM"

6283. .6317

/note="putative"

6335. .6339

6345. .6866

/note="ORF2"

/codon_start=1

/transl_table=11

/protein_id="AAC41426.1"

/db_xref="GI:695280"

/translation="MSGEHLQVYHGDITRMEYDAIVNANGSLGCGGVDAIINGAG
SAIKACRAIRIDTQCGCEPTGEAVITTGHLPAFVYIHAVGVGGDGEDDLANAT
RSIRLAAOHHLRLAFNPISIGIYAFPRERADIAIAAVREALAIAADEIQTIVYVC
IDENRYRVERLS"

6870. .6895

/note="putative"

complement(6870. .6895)

/note="putative"

terminator

terminator

RBS

CDS

misc_signal

-35_signal

-10_signal

promoter

RBS

gene

CDS

PI Falco SC, Cahoon RE, Hiltz WD, Kinney AJ, Rafalski JA;
 XX WPI: 2000-022904/02.
 DR P-PSDB; AAY28439.
 PT Nucleic acid fragments encoding branched chain amino acid biosynthetic
 PR enzymes
 XX
 PS Claim 17; Page 94-95; 102pp; English.

XX AAX89462-X89465 are fragments of corn, soybean, wheat and rice leud
 CC subunit of 3-isopropylmalate dehydratase nucleotide sequences. Sequences
 CC AAX89442-X89465 are nucleic acid fragments that encode all or a
 CC substantial portion of dihydroxyacid dehydratase, a branched chain amino
 CC acid transferase, a leuc or a leud subunit of 3-isopropylmalate
 CC dehydratase from wheat, corn, soybean or rice. These enzymes are involved
 CC in biosynthesis and utilization of branched-chain amino acids. The
 CC nucleic acid sequences can be used to alter the level of expression of a
 CC branched chain amino acid biosynthetic enzyme in a host cell. They can
 CC also be used to obtain a nucleic acid fragment encoding a branched chain
 CC amino acid biosynthetic enzyme, and to identify inhibitors of a branched
 CC chain amino acid biosynthetic enzyme. The encoded proteins may be used
 CC to prepare antibodies for detecting the proteins in situ in cells, or in
 CC vitro in cell extracts.

XX
 SX Sequence 1033 BP; 221 A; 299 C; 282 G; 230 T; 1 other:
 SX
 Query Match 100.0%; Score 1032.6; DB 21; Length 1033;
 Best Local Similarity 100.0%; Pred. No. 2.8e-209;
 Matches 1033; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ATCATGCGCGGCTCTGTGCGGAGCGCGGTGTCACGCGGAGCGGCTTACGCCCAATC 60
 DB 1 ATCATGCGCGGCTCTGTGCGGAGCGCGGTGTCACGCGGAGCGGCTTACGCCCAATC 60
 OY 61 CGAGCTCCACACGAGCGGCTTATCCGCGCTCCAGCTCACTGTCACTCCCTCACTCA 120
 DB 61 CGAGCTCCACACGAGCGGCTTATCCGCGCTCCAGCTCACTGTCACTCCCTCACTCA 120
 OY 121 CTAAATATGCGCGCGCTGCGGCTCATGTCGCCGCGCGCTGTCGCCGCGGCGAGC 180
 DB 121 CTAAATATGCGCGCGCTGCGGCTCATGTCGCCGCGCGCTGTCGCCGCGGCGAGC 180
 OY 181 AGCTGCGCGTGTGAGCGCGTTCCTCAAGGAGAGTCTTCTGAGGCGAGCAATATGAC 240
 DB 181 AGCTGCGCGTGTGAGCGCGTTCCTCAAGGAGAGTCTTCTGAGGCGAGCAATATGAC 240
 OY 241 ACCGACGATCATCCCGCGCGAGCAGCTCACTGTGTCCTTCAACCGCGAGAGTAC 300
 DB 241 ACCGACGATCATCCCGCGCGAGCAGCTCACTGTGTCCTTCAACCGCGAGAGTAC 300
 OY 301 CGCAAGCTGCGTCTTGTGCGTTCGCGGAGGCTCCCATCCGCGGAGCGGCTTC 360
 DB 301 CGCAAGCTGCGTCTTGTGCGTTCGCGGAGGCTCCCATCCGCGGAGCGGCTTC 360
 OY 361 GTGCGTCCGCGGAGAGTCTCCCGCTACGCAATATGTCGCGAGAGCAACTTCGGG 420
 DB 361 GTGCGTCCGCGGAGAGTCTCCCGCTACGCAATATGTCGCGAGAGCAACTTCGGG 420
 OY 421 TGCAGTCTCTCGCGAGCAGCGCGCTGCGCTTGGGCGCGCTGCGAGCGGCAATT 480
 DB 421 TGCAGTCTCTCGCGAGCAGCGCGCTGCGCTTGGGCGCGCTGCGAGCGGCAATT 480
 OY 481 GTTCSGAGAGCTACGCGCGCATCTTTTTCGCAACTCCGCGGAGAGAGGTGTAC 540
 DB 481 GTTCSGAGAGCTACGCGCGCATCTTTTTCGCAACTCCGCGGAGAGAGGTGTAC 540
 OY 541 CCTCTGAGAGCTACGAGAGTGTGGGCTGGAAGAGTCAAGAGAGGAGTGTGTCAC 600
 DB 541 CCTCTGAGAGCTACGAGAGTGTGGGCTGGAAGAGTCAAGAGAGGAGTGTGTCAC 600
 OY 601 GTGAGACTTGTACTCGTCTTTTATTAACACACACTCTGGCAGAGATACAGTGA 660
 DB 601 GTGAGACTTGTACTCGTCTTTTATTAACACACACTCTGGCAGAGATACAGTGA 660

DB 601 GTGAGACTTGTACTCGTCTTTTATTAACACACACTCTGGCAGAGATACAGTGA 660
 OY 661 CCAATGTGATGCTGCGCTGTAATTGAGCGGAGAGATCTTGGCTACGCCGGAG 720
 DB 661 CCAATGTGATGCTGCGCTGTAATTGAGCGGAGAGATCTTGGCTACGCCGGAG 720
 OY 721 ACAGAGATGATGCTGCGGAGAGTCTGCAATGAGGAGAACTTATGAGCGGCTCTG 780
 DB 721 ACAGAGATGATGCTGCGGAGAGTCTGCAATGAGGAGAACTTATGAGCGGCTCTG 780
 OY 781 CGGAGATGAGAGATGAGTGTAGTACGTAAGTACTGCACTTACTGATGTCA 840
 DB 781 CGGAGATGAGAGATGAGTGTAGTACGTAAGTACTGCACTTACTGATGTCA 840
 OY 841 CGGCTCTCAAAATTAAGTGTGCGGCTCCGAAATTAATGATGATCAATTTGGCTT 900
 DB 841 CGGCTCTCAAAATTAAGTGTGCGGCTCCGAAATTAATGATGATCAATTTGGCTT 900
 OY 901 TGTCAAGATGCTTTTGTGTTTGTACTAGTACTGTGCAATTTGATGCTGCTACT 960
 DB 901 TGTCAAGATGCTTTTGTGTTTGTACTAGTACTGTGCAATTTGATGCTGCTACT 960
 OY 961 GTTCTTATCTGTTGATTAATGCTGCTGTCGCAAAATTAATGATGATCAATTTGGCTT 1020
 DB 961 GTTCTTATCTGTTGATTAATGCTGCTGTCGCAAAATTAATGATGATCAATTTGGCTT 1020
 OY 1021 AAAAAAAAAAAAAA 1033
 DB 1021 AAAAAAAAAAAAAA 1033

RESULT 2
 AAX89465
 ID AAX89465 standard; cDNA: 995 BP.
 XX AAX89465;
 AC
 XX
 DT 15-FEB-2000 (first entry)
 XX
 DE Leud subunit of 3-isopropylmalate dehydratase nucleotide sequence.
 XX
 KW Corn; soybean; wheat; rice; dihydroxyacid dehydratase; leuc; leud;
 KW branched chain amino acid transferase; biosynthetic enzyme; antibody;
 KW 3-isopropylmalate dehydratase; ss.
 XX
 OS Trifolium aestivum.
 XX
 FH Key Location/Qualifiers
 FT 1.735
 FT /tag= a
 FT /product= Leud
 XX
 PN W0921880-A2.
 XX
 PD 06-MAY-1999.
 XX
 PF 20-OCT-1998; 98WO-US22081.
 XX
 PR 28-OCT-1997; 97US-0063423.
 XX
 PA (DUPO) DU POINT DE MEMOIRS & CO E I.
 XX
 PI Falco SC, Cahoon RE, Hiltz WD, Kinney AJ, Rafalski JA;
 XX
 DR WPI: 2000-022904/02.
 DR P-PSDB; AAY28442.
 XX
 PT Nucleic acid fragments encoding branched chain amino acid biosynthetic
 PT enzymes
 XX
 PS Claim 17; Page 98; 102pp; English.
 XX
 CC AAX89462-X89465 are fragments of corn, soybean, wheat and rice leud

QY 472 CGCGCATTTGTCGSGAGGGCTACGGCGCATCTTTTTCGCACTCCGTGCGCATGGA 531
DB 528 GCCGAGTGTGCGGAGTCGACGTAGATCTTTCGAACTCGTGCGCACCGGC 587
QY 532 GAGGTGACCTCTGTGAGCTCAGCGACGTTGGGCGCTGGAAGAGTGCACAGAGGAGAT 591
DB 588 GAGGTGATTCGCTAGAG---TCGGAGGAGACGCTCTGAGAGAGTGCACACCGCGCAT 644
QY 592 GTGTGACCGGTGACCTTCTACTACCTCGTTTATTATACACACCTCTGCGAAGAGTAC 651
DB 645 GTGTGACCATTTGATGCTGGAAGAGACCGCTTGATCAATCACACCGGAAGAGATAT 704
QY 652 AACCTGAACCAATTGATGCTGCGCTGTAATTGAGCGGAGGAGATTTTGCTTAC 711
DB 705 CGCTGAACCGATCGGACCGGCGGTCCAGTGTGAGGCGCGGTGCGATCTTTGCCAT 764
QY 712 GCCGGAAGAGCAGATGATGCGTC 737
DB 765 GCCGGAACCGCATGATCCCTC 790

RESULT 5
AAC47222
ID AAC47222 standard; DNA; 945 BP.
XX AAC47222;
AC
XX
XX
DY 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 53012.
XX
KM Hybridisation assay; genetic mapping; gene expression control;
KM protein identification; signal transduction pathway;
KM metabolic pathway; promoter; termination sequence; ss.
XX
OS Arabidopsis thaliana.
XX
PN EF1033405-A2.
XX
PD 06-SEP-2000.
XX
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.

PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 14-JUN-1999; 99US-0138847.
PR 16-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 17-JUN-1999; 99US-0139453.
PR 18-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140693.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144086.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145086.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.

PR 25-FEB-1999; 990S-0121825.
PR 05-MAR-1999; 990S-0123180.
PR 09-MAR-1999; 990S-0123548.
PR 23-MAR-1999; 990S-0125788.
PR 25-MAR-1999; 990S-0126264.
PR 29-MAR-1999; 990S-0126785.
PR 01-APR-1999; 990S-0127462.
PR 06-APR-1999; 990S-0128234.
PR 16-APR-1999; 990S-0128714.
PR 19-APR-1999; 990S-0129845.
PR 21-APR-1999; 990S-0130077.
PR 23-APR-1999; 990S-0130449.
PR 23-APR-1999; 990S-0130510.
PR 23-APR-1999; 990S-0130891.
PR 28-APR-1999; 990S-0131449.
PR 30-APR-1999; 990S-0132048.
PR 30-APR-1999; 990S-0132407.
PR 04-MAY-1999; 990S-0132484.
PR 05-MAY-1999; 990S-0132485.
PR 06-MAY-1999; 990S-0132486.
PR 06-MAY-1999; 990S-0132487.
PR 07-MAY-1999; 990S-0132863.
PR 11-MAY-1999; 990S-0134256.
PR 14-MAY-1999; 990S-0134218.
PR 14-MAY-1999; 990S-0134219.
PR 14-MAY-1999; 990S-0134221.
PR 14-MAY-1999; 990S-0134370.
PR 18-MAY-1999; 990S-0134768.
PR 19-MAY-1999; 990S-0134941.
PR 20-MAY-1999; 990S-0135124.
PR 21-MAY-1999; 990S-0135353.
PR 24-MAY-1999; 990S-0135629.
PR 25-MAY-1999; 990S-0136021.
PR 27-MAY-1999; 990S-0136392.
PR 28-MAY-1999; 990S-0136782.
PR 01-JUN-1999; 990S-0137222.
PR 03-JUN-1999; 990S-0137528.
PR 04-JUN-1999; 990S-0137502.
PR 07-JUN-1999; 990S-0137724.
PR 08-JUN-1999; 990S-0138094.
PR 10-JUN-1999; 990S-0138540.
PR 10-JUN-1999; 990S-0138847.
PR 14-JUN-1999; 990S-0139119.
PR 16-JUN-1999; 990S-0139452.
PR 16-JUN-1999; 990S-0139453.
PR 17-JUN-1999; 990S-0139492.
PR 18-JUN-1999; 990S-0139454.
PR 18-JUN-1999; 990S-0139455.
PR 18-JUN-1999; 990S-0139456.
PR 18-JUN-1999; 990S-0139457.
PR 18-JUN-1999; 990S-0139458.
PR 18-JUN-1999; 990S-0139459.
PR 18-JUN-1999; 990S-0139460.
PR 18-JUN-1999; 990S-0139461.
PR 18-JUN-1999; 990S-0139462.
PR 18-JUN-1999; 990S-0139463.
PR 18-JUN-1999; 990S-0139750.
PR 18-JUN-1999; 990S-0139763.
PR 21-JUN-1999; 990S-0139811.
PR 22-JUN-1999; 990S-0139899.
PR 23-JUN-1999; 990S-0140353.
PR 23-JUN-1999; 990S-0140354.
PR 24-JUN-1999; 990S-0140695.
PR 28-JUN-1999; 990S-0140823.
PR 29-JUN-1999; 990S-0140991.
PR 30-JUN-1999; 990S-0141287.
PR 01-JUL-1999; 990S-0141842.
PR 01-JUL-1999; 990S-0142154.
PR 02-JUL-1999; 990S-0142055.
PR 06-JUL-1999; 990S-0142390.
PR 08-JUL-1999; 990S-0142803.
PR 09-JUL-1999; 990S-0142920.
PR 12-JUL-1999; 990S-0142977.

PR 13-JUL-1999; 990S-0143542.
PR 14-JUL-1999; 990S-0143624.
PR 15-JUL-1999; 990S-0144005.
PR 16-JUL-1999; 990S-0144085.
PR 16-JUL-1999; 990S-0144086.
PR 19-JUL-1999; 990S-0144325.
PR 19-JUL-1999; 990S-0144331.
PR 19-JUL-1999; 990S-0144332.
PR 19-JUL-1999; 990S-0144333.
PR 19-JUL-1999; 990S-0144334.
PR 19-JUL-1999; 990S-0144335.
PR 20-JUL-1999; 990S-0144352.
PR 20-JUL-1999; 990S-0144632.
PR 20-JUL-1999; 990S-0144864.
PR 21-JUL-1999; 990S-0144814.
PR 21-JUL-1999; 990S-0145086.
PR 21-JUL-1999; 990S-0145088.
PR 22-JUL-1999; 990S-0145085.
PR 22-JUL-1999; 990S-0145087.
PR 22-JUL-1999; 990S-0145089.
PR 22-JUL-1999; 990S-0145132.
PR 23-JUL-1999; 990S-0145145.
PR 23-JUL-1999; 990S-0145218.
PR 26-JUL-1999; 990S-0145224.
PR 26-JUL-1999; 990S-0145276.
PR 27-JUL-1999; 990S-0145913.
PR 27-JUL-1999; 990S-0145918.
PR 27-JUL-1999; 990S-0145919.
PR 28-JUL-1999; 990S-0145951.
PR 02-AUG-1999; 990S-0146386.
PR 02-AUG-1999; 990S-0146388.
PR 02-AUG-1999; 990S-0146389.
PR 03-AUG-1999; 990S-0147038.
PR 04-AUG-1999; 990S-0147204.
PR 04-AUG-1999; 990S-0147302.
PR 05-AUG-1999; 990S-0147192.
PR 05-AUG-1999; 990S-0147260.
PR 06-AUG-1999; 990S-0147303.
PR 06-AUG-1999; 990S-0147416.
PR 09-AUG-1999; 990S-0147493.
PR 09-AUG-1999; 990S-0147935.
PR 10-AUG-1999; 990S-0148171.
PR 11-AUG-1999; 990S-0148319.
PR 12-AUG-1999; 990S-0148341.
PR 13-AUG-1999; 990S-0148565.
PR 13-AUG-1999; 990S-0148684.
PR 16-AUG-1999; 990S-0149368.
PR 17-AUG-1999; 990S-0149175.
PR 18-AUG-1999; 990S-0149426.
PR 20-AUG-1999; 990S-0149722.
PR 20-AUG-1999; 990S-0149723.
PR 20-AUG-1999; 990S-0149929.
PR 23-AUG-1999; 990S-0149902.
PR 23-AUG-1999; 990S-0149930.
PR 25-AUG-1999; 990S-0150566.
PR 26-AUG-1999; 990S-0150884.
PR 27-AUG-1999; 990S-0151065.
PR 27-AUG-1999; 990S-0151066.
PR 27-AUG-1999; 990S-0151080.
PR 30-AUG-1999; 990S-0151303.
PR 31-AUG-1999; 990S-0151438.
PR 01-SEP-1999; 990S-0151930.
PR 07-SEP-1999; 990S-0152363.
PR 10-SEP-1999; 990S-0153070.
PR 13-SEP-1999; 990S-0153758.
PR 15-SEP-1999; 990S-0154018.
PR 16-SEP-1999; 990S-0154039.
PR 20-SEP-1999; 990S-0154779.
PR 22-SEP-1999; 990S-0155139.
PR 23-SEP-1999; 990S-0155486.
PR 24-SEP-1999; 990S-0155655.
PR 28-SEP-1999; 990S-0156458.
PR 29-SEP-1999; 990S-0156596.

```

PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 13-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 22-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 26-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

```

Query Match 20.3%; Score 209.2; DB 21; Length 945;

Best Local Similarity 64.9%; Pred. No. 5.7e-35; Indels 12; Gaps 3;

Matches 361; Conservative 0; Mismatches 183;

```

Oy- 202 TTCACGCGGAGTGTCTGCGTGGGCGACAAATATCGACACCGACGATCATCCCGCC 261
Db 238 TTCCATGAGTGTGCTATGTCGTCGGGACAAATCATGACACTGACCAATATCCCGCG 297
Oy 262 GAGCAGCTCACTGTGCTGCTCCCAAGCCGAGAGTACCGCAAGTCCGCTTCCTCCGC 321
Db 298 GAGTTTCTCACTGCGCCCTCGAAATCCAGAGAAATGAGAAAGTGTGATGCTGCAACAT 357
Oy 322 TTCGCGGGGCTCCCATCCGCGGCTTCCGCTCCGCTCCGCTGAGAGATCC 381
Db 358 TTAGTTGGTCTTCCA---GCTTCTTACAAAGAAAGCATTCGTCAGCCAGGTGAGATGAAG 414
Oy 382 TTCGCGTACGCGCATATGTGCGCGGAGCAACTTGGGTCGGTCTTCGCGAGCAC 441
Db 415 ACGAAGTACATCAATCATATGCGCGGTGAAAACCTTGGATGATGCTACGTGAACAT 474
Oy 442 GCGCGCCGTCGCGCTTGGGGCGCGTGGGCGCAAGCCATGTTGSGAGGCTACGCGCGC 501
Db 475 GCTCCGGTGTGTTAGAGAGCAGCGGAGCTAAAGCACTGGTGGCTCATGTATGCTAGA 534
Oy 502 ATCTTTTTCGCACTCCGTGGCCACTGAGAGGCTGTAACCTCGATCGACGAGCATT 561
Db 535 ATCTTTTTCAGAGACTGTTGCTGCTAGTGAAGGTTTATCCTTGGATT---CTGAAGTT 591
Oy 562 GGGGCGCTGAGAGAGTGCAGACAGAGGAGTGTGCTACCGTGGACCTTGTCTACTCCG-- 619
Db 592 AGGGTGTGATGAGTGTACAACTGTGATGTTCCGACTGTTGAGTTGAGGGAAGGAGAT 651
Oy 620 -----TTTTTATTAACCACTCTGGCAAGGATAGACAAGTGAACCAATTTGATGCT 675
Db 652 AGTATTTTGAATCAATCAATACAGACTGGGAAAGATTAAGCTTAAGCCGATTTGGATGCT 711
Oy 676 GGCCCTGTAATTTGAGGCGGAGGAGATCTTTTGCCTACGCCCGGAAGACAGAAATGTCG 735

```

```

Db 712 GGAACCACTGATGCTGCTGTGATATTTGCTATAGTAGAAGAGCTGAATGATCCA 771
Oy 736 TCGAAGCTGCTGCAT 751
Db 772 TCTGCTGCTGCTGAT 787

RESULT 7
ABN98576
ID ABN98576 standard; DNA; 948 BP.
XX
AC ABN98576;
XX
DE 01-AUG-2002 (first entry)
XX
XX Arabidopsis thaliana expressed polynucleotide SEQ ID NO 344.
XX
XX Arabidopsis thaliana; plant; insecticide; fungicide; transgenic; stress;
KW disease; crop; thale cress; tolerance factor; insect; pathogen;
XX nutrition; ds.
XX
XX Arabidopsis thaliana.
XX
XX US2002023281-A1.
XX
PD 21-FEB-2002.
XX
XX 26-JAN-2001; 2001US-0770445.
XX
XX 27-JAN-2000; 2000US-178472P.
XX
XX (GORL/) GORLACH J.
XX (ANYV/) AN Y.
XX (HAMV/) HAMILTON C M.
XX (PRIC/) PRICE J L.
XX (RAIN/) RAINES T M.
XX (YUYV/) YU Y.
XX (RAME/) RAMEKA J G.
XX (PAGE/) PAGE A.
XX (MATH/) MATHEW A V.
XX (LEDF/) LEDFORD B L.
XX (WOES/) WOESSNER J P.
XX (HAAS/) HAAS W D.
XX (GARC/) GARCIA C A.
XX (KRICK/) KRICKER M.
XX (SLAT/) SLATER T.
XX (DAVI/) DAVIS K R.
XX (ALLE/) ALLEN K.
XX (HOEF/) HOFFMAN N.
XX (HURB/) HURBAN P.
XX
XX Gorlach J, An Y, Hamilton CM, Price JL, Raines TM, Yu Y,
XX Rameka JG, Page A, Mathew AV, Ledford BL, Woessner JP, Haas WD;
XX Garcia CA, Kricker M, Slater T, Davis KR, Allen K, Hoffman N;
XX Hurban P;
XX
XX WPI; 2002-400781/43.
XX
XX New Arabidopsis thaliana nucleic acid for identifying homologous genes,
XX producing compositions that modulate the expression or function of its
XX encoded protein, and mapping functional regions of protein
XX
XX Claim 1; SEQ ID NO 344; 49pp + Sequence Listing; English.
XX
XX The invention relates to an Arabidopsis thaliana nucleic acid (I)
XX comprising a sequence capable of hybridizing under stringent conditions
XX to a sequence selected from any one of 999 sequences (ABN98233-ABN99231),
XX given in the specification or its fragment. A polypeptide (II) encoded by
XX (I), a transgenic plant (III) comprising an exogenous nucleic acid or a
XX genetically modified cell (IV) comprising an exogenous nucleic acid, is
XX useful for screening a candidate agent for its biological effect. (I) is
XX useful in identifying homologous or related genes, in producing
XX compositions that modulate the expression or function of its encoded

```

protein, mapping functional regions of the protein and in studying associated physiological pathways. (I) is also useful for the genetic manipulation of cells, particularly plant cells. (I) is also useful in screening assays of various plant strains to determine the strains that are best capable of withstanding a particular disease or environmental stress. (II) and (III) are useful for screening of biologically active agents, e.g. fungicides, insecticides, etc., for elucidating biochemical pathways. The screened agents are useful in improved methods of treating crops to prevent or treat disease. (II) are also useful in screening programs to identify agents that mimic or enhance the action of tolerance factors. Such agents are useful in improved methods of treating crops to enhance their tolerance to environmental stress. (I) is also useful for enhancing or inhibiting production of a biosynthetic product in a plant. (III) is useful for identifying other mediators that may induce expression of proteins of interest, for establishing the extent to which any specific insect and/or pathogen is responsible for damage to a particular plant, for identifying other mediators that enhance or induce tolerance to environmental stress, for identifying factors involved in biosynthetic pathways of nutritional, commercial, or medicinal value and for identifying productions of nutritional, commercial or medicinal value. (IV) is useful in the study of genetic function and regulation, for alteration of the cellular metabolism and for screening compounds that may affect the biological function of the gene or gene products. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from USPTO at seqdata.uspto.gov/sequence.html?docid=999909770445.

Sequence 948 BP; 245 A; 208 G; 287 T; 0 other;

Query Match 20.3%; Score 209.2; DB 24; Length 948;

Best Local Similarity 64.9%; Pred. No. 5,7e-35; Matches 361; Conservative 0; Mismatches 183; Indels 12; Gaps 3;

202 TTCACGCGGAGTCTTGTGTGGGCGACATTCGACACCGCAGATCATCCCGCC 261
 193 TTCACGCGGAGTCTTGTGTGGGCGACATTCGACACCGCAGATCATCCCGCC 252
 262 GAGCACCTACTCTGTGTGGGCGACATTCGACACCGCAGATCATCCCGCC 321
 253 GAGTTTCTACTCTGTGTGGGCGACATTCGACACCGCAGATCATCCCGCC 312
 322 TTCGCGGGGCTCCATCCGCGGCGCTACCCGCGCTTCGCTCCGCGGTGAGAGTCC 381
 313 TTAGTTGGCTTCCA--GCTTCTTCAAGAGACGATTCGCTTCAAGGAGATGATGAG 369
 382 TCCCGCTACGCGCATATGTGTGGGCGACATTCGCGGTGCTCTCCGCGAGCAC 441
 370 AGGAGTACTCATATCATATGCGGTGAGAACTTGTGATGATGATGATGATGATGAT 429
 442 GCGCCGCTGCGCTTGGGCGCGCTGCGCGACGCGCATTTGCGSAGGCGTACGCGCC 501
 430 GCTCGCGGTGTTTGGGAGGACGCGGAGCTTAAAGAGTGGTGGCTTATGCTTGA 489
 502 ATCTTTTTCGCACTCCGCTGGGCGACATTCGCGGTGCTCTCCGCGAGTACGAGT 561
 490 ATCTTTTTCGCACTCCGCTGGGCGACATTCGCGGTGCTCTCCGCGAGTACGAGT 546
 562 GGGGCTGAGAGGATGCAAGACAGGAGATGTGTCACCGTGGACCTTCTACTCCG-- 619
 547 AGGAGTACTCATATCATATGCGGTGAGAACTTGTGATGATGATGATGATGATGAT 606
 620 ----TTTTTATTAACACACCTTGGCAGAGATTAACAGTGAACCAATTTGGATGCT 675
 607 AGATTTTATGATCAATCATACGACGTGGGAGAAAGATTAAGCCGATTTGGATGCT 666
 676 GGGCGCTGATTTGAGCGGAGGAGATCTTTGCTTACGCGCGGAGACAGAGATGATGCG 735
 667 GAGACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 726
 736 TCGAAGAGTGTGATGAT 751
 727 TCGTGTGCTGCTGAT 742

RESULT 8
 ABN98932
 ID ABN98932 standard; DNA: 829 BP.
 AC ABN98932;
 DT 01-AUG-2002 (first entry)
 DE Arabidopsis thaliana expressed polynucleotide seq ID NO 700.
 KW Arabidopsis thaliana; plant; insecticide; fungicide; transgenic; stress;
 KW disease; crop; thale cress; tolerance factor; insect; pathogen;
 KW nutrition; ds.
 OS Arabidopsis thaliana.
 PN US2002023281-A1.
 PD 21-FEB-2002.
 PE 26-JAN-2001; 2001US-0770445.
 PR 27-JAN-2000; 2000US-178472P.
 PA (GORL/) GORLACH J.
 PA (ANY/) AN Y.
 PA (HAM/) HAMILTON C M.
 PA (PRIC/) PRICE J L.
 PA (RAIN/) RAINES T M.
 PA (YUY/) YU Y.
 PA (RAME/) RAMEKA J G.
 PA (PAGE/) PAGE A.
 PA (MATH/) MATHIAS A V.
 PA (LEDF/) LEDFORD B L.
 PA (WOES/) WOESSNER J P.
 PA (HAAS/) HAAS W D.
 PA (GARC/) GARCIA C A.
 PA (KRICK/) KRICKER M.
 PA (SLAT/) SLATER T.
 PA (DAVI/) DAVIS K R.
 PA (ALEX/) ALLEN K.
 PA (HOFF/) HOFFMAN N.
 PA (HURB/) HURBAN P.
 XX
 XX Gorlach J, An Y, Hamilton CM, Price JL, Raines TM, Yu Y;
 XX Rameka JG, Page A, Mathew AV, Ledford BL, Woessner JP, Haas WD;
 XX Garcia CA, Kricker M, Slater T, Davis KR, Allen K, Hoffman N;
 XX Hurban P;
 XX WPI: 2002-400781/43.
 PS
 XX
 XX Claim 1: SEQ ID NO 700; 499p + Sequence Listing: English.
 CC The invention relates to an Arabidopsis thaliana nucleic acid (I)
 CC comprising a sequence capable of hybridizing under stringent conditions
 CC to a sequence selected from any one of 999 sequences (ABN98233-ABN99231),
 CC given in the specification or its fragment. A polypeptide (II) encoded by
 CC (I), a transgenic plant (III) comprising an exogenous nucleic acid or a
 CC genetically modified cell (IV) comprising an exogenous nucleic acid, is
 CC useful for screening a candidate agent for its biological effect. (I) is
 CC useful in identifying homologous or related genes, in producing
 CC compositions that modulate the expression or function of its encoded
 CC protein, mapping functional regions of the protein and in studying
 CC associated physiological pathways. (I) is also useful for the genetic
 CC manipulation of cells, particularly plant cells. (I) is also useful in
 CC screening assays of various plant strains to determine the strains that
 CC are best capable of withstanding a particular disease or environmental
 CC stress. (II) and (III) are useful for screening of biologically active

CC agents, e.g. fungicides, insecticides, etc., for elucidating biochemical
 CC pathways. The screened agents are useful in improved methods of treating
 CC crops to prevent or treat disease. (11) are also useful in screening
 CC programs to identify agents that mimic or enhance the action of tolerance
 CC factors. Such agents are useful in improved methods of treating crops to
 CC enhance their tolerance to environmental stress. (1) is also useful
 CC for enhancing or inhibiting production of a biosynthetic product in a
 CC plant. (111) is useful for identifying other mediators that may induce
 CC expression of proteins of interest, for establishing the extent to which
 CC any specific insect and/or pathogen is responsible for damage to a
 CC particular plant, for identifying other mediators that enhance or induce
 CC tolerance to environmental stress, for identifying factors involved in
 CC biosynthetic pathways of nutritional, commercial, or medicinal value and
 CC for identifying productions of nutritional, commercial or medicinal
 CC value. (IV) is useful in the study of genetic function and regulation,
 CC for alteration of the cellular metabolism and for screening compounds
 CC that may affect the biological function of the gene or gene products.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from USPTO
 CC at seqdata.uspto.gov/sequence.html?docID=99909770445.

XX
 XX Sequence 829 BP; 225 A; 212 C; 182 G; 210 T; 0 other;

Query Match 19.0%; Score 196.6; DB 24; Length 829;

Best Local Similarity 60.3%; Pred. No. 2.6e-32;

Matches 382; Conservative 1; Mismatches 235; Indels 15; Gaps 3;

OY 114 CCACTCACTAAATGCCGCCGCTGGGCTCCATCCCGCGCGGCTCTCCGCGGC 173
 DB 91 CAACCACTACTCTCCCTCCCGCCAGATCAACACGGGTGCTGCATCTCCCGA 150
 OY 174 GGGCAGACGCTCGCGCTGACCGCTTTTCCACGGGAGTCTGCTGGTGGCCACAA 233
 DB 151 TTCAGCGAGATCAATACCAAGAGACTTCCAGCGCTCTGCTCTCTTGAAGCAA 210
 OY 234 TATGACACGACGACGATCATCCCGCGGACACTCTGCTGGTGCCTCCAAAGCCGA 293
 DB 211 CATGACACGATCAATATATCCCGCGGAGTACGACATCTATCCCTTGATTCAGA 270
 OY 294 CGAGTACCGCAAGCTCGCTTCCTTCGCTTCGCGGGCTCCCATCCGCGCTAACCCGAC 353
 DB 271 AGATGCGAGAAACTCGGCTCTTCCGGCTTAACGGCTTACCA--AAATCTCAACGA 327
 OY 354 GCCGTCGCTCGCTCGGCTGAGAGTCTCCCGCTACGCCATCTATGTGGCGGAGCCAA 413
 DB 328 ACGTTTCGTTCTCCAGAGAGATGAATCAAACTCACTCACTCATCTCGCGCGCATYAA 387
 OY 414 CTTGGGGGCGGCTCTCGCGAGACGCGCGCTGCGGCTGGGCGCGTGGCGCAGC 473
 DB 388 TTTGCGTTGGGATCTTCCCGGAACACGCTCCAGTTTGTCTCGGCGCGGCGGAGCTAA 447
 OY 474 CGCCATTGTCGAGGAGCTACGCGCGCATCTTTTTCGCAACTCCGTGGCCACTGGAGA 533
 DB 448 AGCTGTGGTGGCGAATGTACGAGATCTTTTTCGAGAACTGTGAGTACAGGTGA 507
 OY 534 GGTGACCTCTGAGACTCAGGACGTTGGGCGCTCGAAGAGTGCAGAGAGGAGTGT 593
 DB 508 GATTTTCCCGTGTGA--ATCGAGAGTGTAGATTGGCAGCAAAAGAGGGATGT 564
 OY 594 GGTGAC-----CGTGAACCTTGTAACTCGTTTATTAACCAACACTCTGGCAA 644
 DB 565 GGTGACAAATCGAACAAGAGAAAGAGGTAGATTGCTGTCAATCATACAGAGAGAA 624
 OY 645 GAGTACAAAGCTGAACAATTTGATCTGGCCCTGTAATGAGCGGAGGAGATCTT 704
 DB 625 AGAATACAAACTGAAGACCGCTCGGTGATCGCGGTGATGACAGCGCGGTGGAATCTT 684
 OY 705 TGGCTACGCCCGGAGAGACGAGATGTCGCT 737
 DB 685 CGCTTATGCAAGAAAAGCGCGCATGATTCCTTC 717

RESULT 9

AAAT70724
 ID AAAT70724 standard; DNA; 4347 BP.
 XX
 XX
 AC AAAT70724;
 XX
 DT 15-DEC-2000 (first entry)
 XX
 DE Thermus thermophilus homoacetylase hydratase gene region.
 XX
 KW Genome; L-lysine biosynthesis; homoacetylase hydratase; ds.
 XX
 OS Thermus thermophilus.
 XX
 PN JP2000157276-A.
 XX
 PD 13-JUN-2000.
 XX
 PF 24-NOV-1998; 98JP-0333132.
 XX
 PR 24-NOV-1998; 98JP-0333132.
 XX
 PA (AJIN) AJINOMOTO KK.
 XX
 DR WPI: 2000-485355/43.
 XX
 DR P-PSDB; AAB15394, AAB15395.
 XX
 PT L-lysine biosynthetic system gene of Thermus genus microbe -
 XX
 PS Claim 4; Page 10-12; 14pp; Japanese.
 XX

This sequence represents the genomic DNA region from Thermus thermophilus
 CC which contains the coding regions for the L-lysine biosynthetic protein
 CC homoacetylase hydratase subunits I and II. The gene can be used for the
 CC generation of L-lysine-producing Thermus thermophilus strain for the
 CC production of L-lysine.

XX
 XX Sequence 4347 BP; 680 A; 1423 C; 1569 G; 675 T; 0 other;

Query Match 6.3%; Score 65.2; DB 21; Length 4347;

Best Local Similarity 55.4%; Pred. No. 0.00025;

Matches 124; Conservative 1; Mismatches 99; Indels 0; Gaps 0;

OY 374 AGAGATCTCCCGCTACGACATATGTCGCGGAGCAACTTCGGTGGGCTCTCTC 433
 DB 1857 AGGAGTGGGCGCCCGGAGATCTCTGCTTCGCGGAGCAACGCGGCTCGCACCC 1916
 OY 434 GCGAGCAGCGCGCGCTTGGGCGCTGGGCGCACGCGCATTTGTGCGSAGGGCT 493
 DB 1917 GCGAGTACGCGCCCGGAGCGCTGAAGCGCTCGGCTCGGCGCATATCGCAAAAGCT 1976
 OY 494 ACCGCGCATCTTTTTCGCACTCCGTCGCGACGTGAGAGGTAACTCTGAGACTCA 553
 DB 1977 ACCCGCGCATCTTTTTCGCACTCCGTCGAGTACCTGGGATCTCCCTTTGATGGAGAG 2036
 OY 554 CGGACGTTGGGCGCTGGAAGAGTGCAGAGAGGAGGATGTGTC 597
 DB 2037 AGGTGTGATGTGCTAGAGAGAGCGGACGAGGTGAGTGTGAC 2080

RESULT 10

AAAT29037

ID AAT29037 standard; cDNA; 217 BP.

XX
 XX AAT29037;

XX
 XX 21-NOV-2001 (first entry)

XX
 XX cDNA encoding for human DNA-binding protein #8.

XX
 XX Human; DNA-binding protein; histone; chromo domain protein;
 KW chromatin organisation modifier; Y-box binding protein;
 KW DNA organisation; gene transcription; malignant disease;
 KW autoimmune disorder; rheumatic disease; genetic abnormality;

KW infectious disease; neurological disorder; gene therapy;
KW immunomodulatory; anti-HIV; anti rheumatic; anti microbial;
KW cytostatic; ss.
XX
OS Homo sapiens.
XX
PN WO200155162-A1.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01305.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214866.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216860.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 23-AUG-2000; 2000US-0227182.
PR 30-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0228287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231444.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0233080.
PR 08-SEP-2000; 2000US-0233081.
PR 12-SEP-2000; 2000US-0231966.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.

PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246611.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.

PR	29-SEP-2000;	2000US-0236368.
PR	29-SEP-2000;	2000US-0236369.
PR	29-SEP-2000;	2000US-0236370.
PR	02-OCT-2000;	2000US-0236802.
PR	02-OCT-2000;	2000US-0237037.
PR	02-OCT-2000;	2000US-0237038.
PR	02-OCT-2000;	2000US-0237039.
PR	02-OCT-2000;	2000US-0237040.
PR	13-OCT-2000;	2000US-0239935.
PR	13-OCT-2000;	2000US-0239937.
PR	20-OCT-2000;	2000US-0240960.
PR	20-OCT-2000;	2000US-0241221.
PR	20-OCT-2000;	2000US-0241785.
PR	20-OCT-2000;	2000US-0241786.
PR	20-OCT-2000;	2000US-0241787.
PR	20-OCT-2000;	2000US-0241808.
PR	20-OCT-2000;	2000US-0241809.
PR	01-NOV-2000;	2000US-0244517.
PR	08-NOV-2000;	2000US-0246474.
PR	08-NOV-2000;	2000US-0246475.
PR	08-NOV-2000;	2000US-0246476.
PR	08-NOV-2000;	2000US-0246477.
PR	08-NOV-2000;	2000US-0246478.
PR	08-NOV-2000;	2000US-0246523.
PR	08-NOV-2000;	2000US-0246524.
PR	08-NOV-2000;	2000US-0246525.
PR	08-NOV-2000;	2000US-0246526.
PR	08-NOV-2000;	2000US-0246527.
PR	08-NOV-2000;	2000US-0246528.
PR	08-NOV-2000;	2000US-0246532.
PR	08-NOV-2000;	2000US-0246609.
PR	08-NOV-2000;	2000US-0246610.
PR	08-NOV-2000;	2000US-0246611.
PR	17-NOV-2000;	2000US-0249207.
PR	17-NOV-2000;	2000US-0249208.
PR	17-NOV-2000;	2000US-0249209.
PR	17-NOV-2000;	2000US-0249210.
PR	17-NOV-2000;	2000US-0249211.
PR	17-NOV-2000;	2000US-0249212.
PR	17-NOV-2000;	2000US-0249213.
PR	17-NOV-2000;	2000US-0249214.
PR	17-NOV-2000;	2000US-0249215.
PR	17-NOV-2000;	2000US-0249216.
PR	17-NOV-2000;	2000US-0249217.
PR	17-NOV-2000;	2000US-0249218.
PR	17-NOV-2000;	2000US-0249244.
PR	17-NOV-2000;	2000US-0249245.
PR	17-NOV-2000;	2000US-0249264.
PR	17-NOV-2000;	2000US-0249265.
PR	17-NOV-2000;	2000US-0249297.
PR	17-NOV-2000;	2000US-0249299.
PR	17-NOV-2000;	2000US-0249300.
PR	01-DEC-2000;	2000US-0250160.
PR	01-DEC-2000;	2000US-0250391.
PR	05-DEC-2000;	2000US-0251030.
PR	05-DEC-2000;	2000US-0251988.
PR	05-DEC-2000;	2000US-0256719.
PR	06-DEC-2000;	2000US-0251479.
PR	08-DEC-2000;	2000US-0251866.
PR	08-DEC-2000;	2000US-0251868.
PR	08-DEC-2000;	2000US-0251869.
PR	08-DEC-2000;	2000US-0251989.
PR	11-DEC-2000;	2000US-0251990.
PR	05-JAN-2001;	2001US-0259678.
XX		
PA	(HUMA-)	HUMAN GENOME SCI INC.
XX		
PI	Rosen CA, Barash SC, Ruben SM,	
XX	WPI: 2001-465557/50.	

DR	P-PSDB; AAU18238.
XX	Nucleic acid molecules encoding human secreted chromosomal binding
PT	proteins, used in preventing, treating or ameliorating a disorder, e.g.
PT	Alzheimer's and Parkinson's diseases and cancers -
XX	
PS	Claim 4; SEQ ID NO 95; 561bp; English.
XX	
CC	The present invention relates to the isolation of novel DNA-binding
CC	proteins (AAU18154-AAU18281), and cDNA and genomic sequences encoding
CC	for these proteins. DNA-binding proteins such as histones, chromo
CC	(chromatin organisation modifier) domain proteins, and v-box binding
CC	proteins may contribute to diseases resulting from aberrant DNA
CC	organisation and/or gene transcription. The sequences of the invention
CC	are useful in screening assays to identify antagonists and/or agonists
CC	that may enhance or block activities mediated by DNA-binding proteins.
CC	Blocks of DNA-binding proteins may be useful in treating disorders
CC	such as malignant diseases (e.g. cancer), autoimmune disorders
CC	(e.g. diabetes mellitus), rheumatic diseases (e.g. rheumatoid
CC	arthritis), genetic abnormalities (e.g. cystic fibrosis), infectious
CC	diseases (e.g. HIV) and neurological disorders (e.g. Alzheimer's
CC	disease). The polynucleotide sequences of the invention may also be
CC	used in gene therapy. AA529030-AA529157 represent cDNA sequences
CC	encoding for novel DNA-binding proteins.
CC	Note: The sequence data for this patent did not form part of the printed
CC	specification, but was obtained in electronic format directly from WIPO
CC	at ftp://wipo.int/pub/published_pct_sequences .
XX	
SQ	Sequence 297 bp; 129 A; 47 C; 51 G; 68 T; 2 other;
	Query Match 6.2%; Score 63.6; DB 22; Length 297;
	Best Local Similarity 71.2%; Pred. No. 0.00029;
	Matches 84; Conservative 0; Mismatches 34; Indels 0; Gaps
QY	916 TTTTGTGTACTACTTGTGACAAATTGTACTCTGCTGCTACTGTCTTATCTGTTTG 975
DB	148 TCTTTTGTATTAATTAATAGCTTAACTGTAGTGTCTTAGTAGTGACATTTCTGTCTCTC 207
QY	976 AATACTGCTCTGTGTGGCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1033
DB	208 AATAAATTTTACTTGTCTGCACAAAAAATAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 265
	RESULT 12
ID	AA189088
AC	AA189088 standard; cDNA; 387 bp.
XX	
AC	AA189088;
XX	
DT	06-NOV-2001 (first entry)
XX	
DE	Human polynucleotide SEQ ID NO 9148.
XX	
RW	Human; cytokine; cell proliferation; cell differentiation; gene therapy;
RW	vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
RW	tissue growth factor; immunomodulatory; cancer; leukaemia;
RW	nervous system disorders; arthritis; inflammation; ss.
XX	
OS	Homo sapiens.
XX	
FN	WO200164835-A2.
XX	
PD	07-SEP-2001.
XX	
PF	26-FEB-2001; 2001WO-US04927.
XX	
PR	28-FEB-2000; 2000US-0515126.
PR	18-MAY-2000; 2000US-0577409.
XX	
PA	(HYSE-) HYSEQ INC.
XX	
PI	Tang YT, Liu C, Drmanac RT;
XX	


```

XX 04-DEC-1997; 97US-0984919.
PF
XX
XX 03-DEC-1996; 96US-0781420.
PR
XX 12-JUN-1997; 97US-0874102.
XX
XX (HESK-) HESKA CORP.
XX
XX Chandrashekar R;
XX
XX WPI; 2002-451385/48.
XX
XX New nucleic acid encoding nematode transglutaminase, useful in vaccines
PT for treating or preventing nematode infestation in humans and animals
XX
XX
XX Example 7; Column 61-63; 66pp; English.
XX
XX The invention relates to an isolated nucleic acid (I) encoding nematode
CC transglutaminase. (I), optionally incorporated into recombinant viruses
CC or cells, are used to treat or prevent infestation by parasitic nematodes
CC in humans or animals, especially by Dirofilaria immitis, Brugia malayi
CC or Onchocerca volvulus. (I) can also be used for expression of the
CC corresponding recombinant proteins which are also useful in
CC antineoplastic vaccines and for raising antibodies (useful for treatment
CC or diagnosis). ABR82304-ABR82347 represent nematode transglutaminase
CC coding sequences and primers of the invention.
XX
XX Sequence 1466 BP; 520 A; 209 C; 298 G; 439 T; 0 other;
SQ
XX
XX Query Match 5.5%; Score 57; DB 24; Length 1466;
XX Best Local Similarity 66.9%; Pred. No. 0.011;
XX Matches 81; Conservative 0; Mismatches 40; Indels 0; Gaps 0;
XX
XX QY 913 TTTTGTGTTACTGCTGTGCAATTGTGACTCTGCTGCTGCTGCTGTTACTGT 972
XX | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX Db 1345 TTTTGTGTTACTGCTGTGCAATTGTGACTCTGCTGCTGCTGCTGTTAAAT 1404
XX
XX QY 973 TTGAATTAAGTCTGCTGTGCCAAAAAATAAAAAAAAAAAAAAAAAAAAAA 1032
XX | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX Db 1405 TTTAAATTAATTTTGTGAAAAAATAAAAAAAAAAAAAAAAAAAAAA 1464
XX
XX QY 1033 A 1033
XX
XX Db 1465 A 1465
XX
XX RESULT 15
XX ABR82309/C
XX ID ABR82309 standard; cDNA; 1466 BP.
XX
XX AC ABR82309;
XX
XX 27-AUG-2002 (first entry)
XX
XX DE Dirofilaria immitis transglutaminase gene #6.
XX
XX KW Nematode; transglutaminase; nematocid; Dirofilaria immitis;
XX Brugia malayi; Onchocerca volvulus; antineoplastic vaccine; gene; ss.
XX
XX Dirofilaria immitis.
XX
XX OS
XX
XX PN US6383774-B1.
XX
XX PD 07-MAY-2002.
XX
XX PF 04-DEC-1997; 97US-0984919.
XX
XX PR 03-DEC-1996; 96US-0781420.
XX
XX PR 12-JUN-1997; 97US-0874102.
XX
XX PA (HESK-) HESKA CORP.
XX

```

```

PI Chandrashekar R;
XX
XX WPI; 2002-451385/48.
XX
XX New nucleic acid encoding nematode transglutaminase, useful in vaccines
PT for treating or preventing nematode infestation in humans and animals
XX
XX
XX Example 7; Column 67; 66pp; English.
XX
XX The invention relates to an isolated nucleic acid (I) encoding nematode
CC transglutaminase. (I), optionally incorporated into recombinant viruses
CC or cells, are used to treat or prevent infestation by parasitic nematodes
CC in humans or animals, especially by Dirofilaria immitis, Brugia malayi
CC or Onchocerca volvulus. (I) can also be used for expression of the
CC corresponding recombinant proteins which are also useful in
CC antineoplastic vaccines and for raising antibodies (useful for treatment
CC or diagnosis). ABR82304-ABR82347 represent nematode transglutaminase
CC coding sequences and primers of the invention.
XX
XX Sequence 1466 BP; 440 A; 297 C; 209 G; 520 T; 0 other;
SQ
XX
XX Query Match 5.5%; Score 57; DB 24; Length 1466;
XX Best Local Similarity 66.9%; Pred. No. 0.011;
XX Matches 81; Conservative 0; Mismatches 40; Indels 0; Gaps 0;
XX
XX QY 913 TTTTGTGTTACTGCTGTGCAATTGTGACTCTGCTGCTGCTGCTGTTACTGT 972
XX | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX Db 122 TTTTGTGTTACTGCTGTGCAATTGTGACTCTGCTGCTGCTGCTGTTAAAT 63
XX
XX QY 973 TTGAATTAAGTCTGCTGTGCCAAAAAATAAAAAAAAAAAAAAAAAAAAAA 1032
XX | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX Db 62 TTTAAATTAATTTTGTGAAAAAATAAAAAAAAAAAAAAAAAAAAAA 3
XX
XX QY 1033 A 1033
XX
XX Db 2 A 2
XX
XX Search completed: April 3, 2003, 03:23:08
XX Job time : 1585 secs

```


Db 181 AGTCGCCGTGTCAGCGGCTTTTCCAGCGGAGCTCTGCTGCTGGGCGCAATATCGAC 240
QY 241 ACCGACGATATCCCGCCGAGACCTACTGTGTCCCTCCAAAGCCGAGAGTAC 300
Db 241 ACCGACGATATCCCGCCGAGACCTACTGTGTGTCCCTCCAAAGCCGAGAGTAC 300
QY 301 CGCAGCTGCTTCTGCTTGTGCGGGGCTCCATCCGCGGCTACCGGAGCGCTTC 360
Db 301 CGCAGCTGCTTCTGCTTGTGCGGGGCTCCATCCGCGGCTACCGGAGCGCTTC 360
QY 361 GTCCGCTCCGCTGAGAGAGTCTCCGCTACGCTATTTTCGCGGAGCCAACTTCGGG 420
Db 361 GTCCGCTCCGCTGAGAGAGTCTCCGCTACGCTATTTTCGCGGAGCCAACTTCGGG 420
QY 421 TCGCGTCTCTCGGAGACGCGCCGCTGCGGCTTGGGCGCGCTGCGGAGCGCTAT 480
Db 421 TCGCGTCTCTCGGAGACGCGCCGCTGCGGCTTGGGCGCGCTGCGGAGCGCTAT 480
QY 481 GTTCCGAGGCTACGCGGCTATTTTTCGCACTCTGCGGCTGCGGAGAGTGTAC 540
Db 481 GTTCCGAGGCTACGCGGCTATTTTTCGCACTCTGCGGCTGCGGAGAGTGTAC 540
QY 541 CCTCGAGCTACGAGAGTGGGGCTGGAGAGAGTGCAGAGAGGAGTGTGTAC 600
Db 541 CCTCGAGCTACGAGAGTGGGGCTGGAGAGAGTGCAGAGAGGAGTGTGTAC 600
QY 601 GTGACCTTGTCTAATCCGTTTATTAACCACTCTGCGGAGAGTACAAAGCTGAAA 660
Db 601 GTGACCTTGTCTAATCCGTTTATTAACCACTCTGCGGAGAGTACAAAGCTGAAA 660
QY 661 CCAATGGTGTGCTGCGGCTGTAAATGAGCGGAGAGAGTGTGCTGCGGAGAG 720
Db 661 CCAATGGTGTGCTGCGGCTGTAAATGAGCGGAGAGAGTGTGCTGCGGAGAG 720
QY 721 ACAGGATGATTCGCTGAGAAAGTGTGATGAGAGAGAGCTTATGACGCGGCTTC 780
Db 721 ACAGGATGATTCGCTGAGAAAGTGTGATGAGAGAGAGCTTATGACGCGGCTTC 780
QY 781 CGGAGATGAGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 840
Db 781 CGGAGATGAGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 840
QY 841 CGGATGCTCAAAATAGTGGGCTACCGAAATATGATGATGATGATGATGATGATG 900
Db 841 CGGATGCTCAAAATAGTGGGCTACCGAAATATGATGATGATGATGATGATGATG 900
QY 901 TGTCAAGATGCTTTTGTGTACTAGTGTGATGATGATGATGATGATGATGATGATG 960
Db 901 TGTCAAGATGCTTTTGTGTACTAGTGTGATGATGATGATGATGATGATGATGATG 960
QY 961 GTTCTATCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1020
Db 961 GTTCTATCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1020
QY 1021 AAAAAAAAAA 1033
Db 1021 AAAAAAAAAA 1033

RESULT 2

US-10-027-450-52

; Sequence 52, Application US/10027450

; Patent No. US20020102715A1

; GENERAL INFORMATION:

; APPLICANT: Faico, Saverio Carl

; APPLICANT: Hitz, William D.

; APPLICANT: Kinney, Anthony J.

; APPLICANT: Cahoon, Rebecca E.

; APPLICANT: Rafalski, J. Antoni

; TITLE OF INVENTION: PLANT BRANCHED CHAIN AMINO ACID BIOSYNTHETIC ENZYMES

; FILE REFERENCE: BB-1126

; CURRENT APPLICATION NUMBER: US/10/027,450

; CURRENT FILING DATE: 2001-12-20

;; PRIOR APPLICATION NUMBER: 60/063,423
;; PRIOR FILING DATE: 1997 October 28
;; NUMBER OF SEQ ID NOS: 34
;; SOFTWARE: Microsoft Word Version 7.0A
;; SEQ ID NO 52
;; LENGTH: 995
;; TYPE: DNA
;; ORGANISM: Trilicium aestivum
US-10-027-450-52

Query Match 93.0%; Score 961.2; DB 12; Length 995;
Best Local Similarity 99.0%; Pred. No. 1,8e-255;
Matches 978; Conservative 1; Mismatches 4; Indels 5; Gaps 1;

QY 26 CGGGGATGTCACGAGCGCTTCTAGCCCATTCGAGCTCCAAACGAGCGGTTATCC 85
Db 8 CGGGGATGTCACGAGCGCTTCTAGCCCATTCGAGCTCCAAACGAGCGGTTATCC 67

QY 86 GGGCTCCCACTACCTGTATCGCTCCCTACCTACCTAAATGCGCGCTGGTCCA 145
Db 68 GGGCTCCCACTACCTGTATCGCTCCCTACCTACCTAAATGCGCGCTGGTCCA 127

QY 146 TCGTCCCGGCGGCGCTGCTGCGGCGGCGGCGAGCTGCGGCTGCGGCTGCTTCC 205
Db 128 TCGTCCCGGCGGCGCTGCTGCGGCGGCGGCGAGCTGCGGCTGCGGCTGCTTCC 187

QY 206 ACGGCGATGCTTGTGTGTGCGGCGAGATGTGACACCGACCGAGATCATCCCGCGAC 265
Db 188 ACGGCGATGCTTGTGTGTGCGGCGAGATGTGACACCGAGATCATCCCGCGAGAC 247

QY 286 ACGTACCTGCTGCTCCCTCCAAAGCGGAGAGTACCGCAAGCTGCTGCTTCCGCTTCG 325
Db 248 ACGTACCTGCTGCTCCCTCCAAAGCGGAGAGTACCGCAAGCTGCTGCTTCCGCTTCG 307

QY 326 CGGGGCTCCCACTCGCGGCTTACCGGAGCGCTTGTGCTGCTGCTGCTGCTGCTGCT 385
Db 308 CGGGGCTCCCACTCGCGGCTTACCGGAGCGGCTTGTGCTGCTGCTGCTGCTGCTGCT 367

QY 386 GCTACGCGCATATGTGTGCGGCGAGCGCACTTGGGCTGCTGCTTCCGCGAGCGCGC 445
Db 368 GCTACGCGCATATGTGTGCGGCGAGCGCACTTGGGCTGCTGCTTCCGCGAGCGCGC 427

QY 446 CCGTCCGCTTGGGCGCGCTGCGGCGAGCGCATTTGTGCGGAGGCTTACCGGCGCATCT 505
Db 428 CCGTCCGCTTGGGCGCGCTGCGGCGAGCGCATTTGTGCGGAGGCTTACCGGCGCATCT 487

QY 506 TTTTTCGCACTCGTGGCGCACTGAGAGGTGTACCTGCGAGCTGCGAGCGTGGGG 565
Db 488 TTTTTCGCACTCGTGGCGCACTGAGAGGTGTACCTGCGAGCTGCGAGCGTGGGG 547

QY 566 CCGGAGAGGTGCAAGACAGGGGATGTGTCAACCGTGTGCACTCCGTTTAA 625
Db 548 CCGGAGAGGTGCAAGACAGGGGATGTGTCAACCGTGTGCACTCCGTTTAA 607

QY 626 TTAACGACACTCGGCAAGAGTGTGCAAGCTGAACCAATGCTGCTGCGGCGCTGTA 685
Db 608 TTAACGACACTCGGCAAGAGTGTGCAAGCTGAACCAATGCTGCTGCGGCGCTGTA 667

QY 686 TTTAGGCGGAGGATCTTGTGCTACGCGCGGAGAGAGAGATGATGCTGGAAGCTG 745
Db 668 TTTAGGCGGAGGATCTTGTGCTACGCGCGGAGAGAGAGATGATGCTGGAAGCTG 727

QY 746 CTGCTATGAGGAA-----ACTTATGCGAGCGGCTTCCGAGAGTGAAGAGTAAAGT 800
Db 728 CTGCTATGAGGAAAGATGATGATGAGCGGAGCTTCCGAGAGTGAAGAGTAAAGT 787

QY 801 GGAGTTAGAGTAAAGTACTGCACTTATGATGATGATGATGATGATGATGATGATG 860
Db 788 GGAGTTAGAGTAAAGTACTGCACTTATGATGATGATGATGATGATGATGATGATG 847

QY 861 CGGCTACCGAAATATGATGATGATGATGATGATGATGATGATGATGATGATGATG 920
Db 848 CGGCTACCGAAATATGATGATGATGATGATGATGATGATGATGATGATGATGATG 907

[illegible]

	RESULT 3	
	US-10-027-450-48	
	: Sequence 48, Application US/10027450	
	: Patent No. US20020102715A1	
	: GENERAL INFORMATION:	
	: APPLICANT: Falco, Saverio Carl	
	: APPLICANT: Hitz, William D.	
	: APPLICANT: Kinney, Anthony J.	
	: APPLICANT: Cahoon, Rebecca E.	
	: APPLICANT: Rafalski, J. Antoni	
	: TITLE OF INVENTION: PLANT BRANCHED CHAIN AMINO ACID BIOSYNTHETIC ENZYMES	
	: FILE REFERENCE: BB-1126	
	: CURRENT APPLICATION NUMBER: US/10/027,450	
	: CURRENT FILING DATE: 2001-12-20	
	: PRIOR APPLICATION NUMBER: 60/065,423	
	: PRIOR FILING DATE: 1997 October 28	
	: NUMBER OF SEQ ID NOS: 54	
	: SOFTWARE: Microsoft Word Version 7.0A	
	: SEQ ID NO 48	
	: * LENGTH: 1112	
	: TYPE: DNA	
	: ORGANISM: Oryza sativa	
	US-10-027-450-48	
-Query Match	44.7%; Score 462.2; DB 12; Length 1112;	
Best Local Similarity	78.2%; Pred. No. 1.3e-117;	
Matches 570; Conservative 1; Mismatches 149; Indels 9; Gaps 1,		
OY	27 GGCGGATGCCACGGCAGCGCTTCTAGCCCAATCCGAGCTCACACGACGCCTTTATTCCG 86	
Dd		
OY	72 GGGGGGCGCGGTGACAGACTTGTCGGACCAGGTGCCACGGCCCTCGAGAGAGCTTCGCCGC 131	
OY	87 GCCTGCCAGCTACCTGTGATCGCTTCACACTACTAAATGCCGCCGCGCTGGTCCAT 146	
Dd		
OY	132 CCGACAGCTGGGTGCGCGCTATCTGCGCGGCCCGCCCTGTAAAGCCACACACAGTGTCCCT 191	
OY	147 C-----GTCCCSCGGCCGTCTGTCTGCGCGCGGGGAGCATCTGCGCTGCTGACG 197	
Dd		
OY	192 GAOCGCGGTGTGCGCGCGGTGCGGCTGCGGCTGCGCGGGGGAGCTGAAGCTGCGCGCG 251	
OY	198 CGTTTTCAAGCGGATGCTGTGTGTGGTGGCGACAATGTGACACGACCAAGATCATCC 257	
Dd		
OY	252 CGATTTCAGGGGAAGTGTCTGTGTGGGGATTAACATGACACCGAACCAAGATCATCC 311	
OY	258 CGCGAGACCTTACTCTGTGTGCTCCAGAGCGGACGAGTAGACGCAAGCTCGATTCTT 317	
Dd		
OY	312 GCGCGAGACACTATCCCTGTCCCTCCAAGCCGACAGTAGACCGCAAAGCTCGGCTGTT 371	
OY	318 CGCCTTCGGGGGCTCCATCCGCGGAGCGTACCGACGCGCGCTGTGCGCTCCGGGTAGGA 377	
Dd		
OY	372 CGCCTTCGTGTGCTCTCCCAACCGCGGCTTACCCAGCGCTTCTGTGCGCCCCCGAGAGA 431	
OY	378 GTCTCCCGCTACGCGCATATTGTGCGGGAGCCAACTTGGGTTGCGGTTCTCTGCGGA 437	
Dd		
OY	432 GACACACCGCTACGCGCTCATCATGTGCGGGCGCCAACTTCGGGTGCGGCTCTCCGCGA 491	
OY	438 GCAAGCGCCGCTGCGCTTGGGGCGCGTGGCGACAAGCCCACTTTCGCSAGGGGCTACGG 497	
Dd		
OY	492 GCAAGCGCCGCTGCGCTTGGGGCGCGCGCGCGCCGCCGCCGCTGTGTGGCCAGAGGCTACGC 551	
OY	498 GCGCATCTTTTTCGCAATCTCGTGCGCACTGAGAGAGGTGTAACCTCTGTGAGCTCAACGA 557	
Dd		
OY	552 GCGCATCTTTTTCGCAATCTCGTGCGCACTGAGAGAGGTGTAACCTCTGTGAGCTCAACGA 611	

QY	558	CGTTGGGGCCCTGGAGGAGTGCAGACAGGGGATGTGTGTCACCGTGGACCTTGTCTAACTC	617
Db	612	CACGTGAGGCTCGAAGGAGTCCAAAGACCGGGGATGTGTGTCACGGTGGAACTTGATTAATTG	671
QY	618	CGTTTTTAAATCAACACCTCTGGCTCAAGGAGATCAAGCTGAACCAATTTGGTGTATGCTGG	677
Db	672	CGTCATGATCAACACACATCTCGGCAAGACAGTACCAAGCTGTAAAGCTATTCGGCGATGCCGG	731
QY	678	CCCTGTAAATTGAAGGCGGAGGGAATCTTTGGCTACGGCCGGGAAGACAGGAAATGATTGCTC	737
Db	732	GCCGGTTAATTGAAGGACGCGGGATCTTTGCCCTATGCCCGGAAGACCGGATGATGATGC	791
QY	738	GAAAGCTGC	746
Db	792	CAAGTCTGC	800

[illegible]


```
OY 532 GAGGTGTACCTCTGGAGCTACGAGCTTGGGGCTTGAAGAGATGCAAGACAGGGAT 591
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 588 GAGGTGTATCCCTTAGAG---TCGAGGGAGCCCTCTGCGAGAGATGACACACCGGAT 644
OY 592 GTGTGACCGGTGAGCTTGTCTAATCCTGTTTATTAACACACCTTGGCAAGAGATAC 651
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 645 GTGTGACGATTTAGAGCTGAGAGAGAGCCGCTGATCAATCAACACCGAAGAGATAT 704
OY 652 AAGCTAAACCAATGTGTATGCTGCGCCCTGTAATTTGAGGGGGAGGATCTTTCCTAC 711
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 705 CGCTTAACCAACCGATGCGGACGCGGTCCAGATGAGGCGGATGACCTTTCTCTAT 764
OY 712 GCCCGAGACAGAGATGATGCTG 737
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 765 GCCAGGAAACCGGATGATTCCTC 790
```

RESULT 5

```
US-09-938-842A-996
; Sequence 996, Application US/09938842A
; Patent No. US20020160378A1
; GENERAL INFORMATION:
; APPLICANT: Harper, Jeff
; APPLICANT: Kieps, Joel
; APPLICANT: Wang, Xun
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
; TITLE OF INVENTION: SAME, AND METHODS OF USE
; FILE REFERENCE: SCRIPI300-3
; CURRENT APPLICATION NUMBER: US/09/938,842A
; PRIOR FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: US 60/227,866
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: US 60/264,647
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/300,111
; PRIOR FILING DATE: 2001-06-22
; NUMBER OF SEQ ID NOS: 5379
; SEQ ID NO 996
; LENGTH: 756
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-938-842A-996
```

Query Match 20.3%; Score 210; DB 9; Length 756;
Best Local Similarity 65.2%; Pred. No. 4.5e-48;
Matches 360; Conservative 0; Mismatches 180; Indels 12; Gaps 3;

```
OY 202 TTCACGGCGAGTCTGTCGTGGGGGAGCAATTCGACACCGACAGATATCCCGCC 261
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 208 TTCATGAGACTGTCTATGTGTGCGGACACATCTGCACACTGACCAATCATTCGCCG 267
OY 262 GAGCACTCACTCTGTGCTCCCTCCAGCCGAGACGATACCGCAAGCTGCTCTTCCG 321
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 268 GAGTTTCACTCTGCTCCCTTCGATCCAGAGAAATACGAAATCGGTTCTTACGT 327
OY 322 TTCGGGGGCTCCATCGCGGCTTACCGACGCGCTTCTGCGCTCCGGGGAGAGATCC 381
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 328 TTAGTTGCTTCCA--GCTTTTACAAAGGAGATTCGTCACCGAGGAGATAG 384
OY 382 TCCCGCTACGCAATATGTGCGGAGACCAATTCGAGTGCGGTTCTCTCGGAGAC 441
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 385 ACGAAGTCACTATCATATATGCGGGTGAACACTTTGATGTGATCGTCACTGAAAT 444
OY 442 GCGCCGTCGCGCTTGGGGCGCTGCGCACCGCCCATTTGTCGAGGGCTACGCGCG 501
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 445 GCTCGGTTTGTATTAGGAGACGCGGAGCTAAAGCATGTGCTCACTTATGCTAGA 504
OY 502 ATCTTTTTCGAATCTCGTGGCCACTGAGAGGCTGTAACCTTGAGAGCTCAGGAGCT 561
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 505 ATCTTTTTCGAATCTCTGTCTACTGTGTAGGTTTATTCCTTGTGATTT---CTGAG 561
```

```
OY 562 GGGGCTGGAAGAGTGCACAGACAGGGGATGTGTACCCGTGACCTTCTACTCG-- 619
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 562 AGGGTTTGTATGATGTGTACTACCTGGTGTATGCTTCCAGCTTTAGTTGAGGAGAGAT 621
OY 620 ----TTTTTAACACACACCTCTGGCAGAGATACAAAGCTGAACCAATTGGTATGCT 675
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 622 AGTATTTTGCATCATCATACATGACCTGGGAAAGAGTACAAAGCTTAAGCCGATGGTATGCT 681
OY 676 GGCCCTTAATTTAGGCGGGAGAGGATTTTTCCTACGCCCCGGGAAGACAGATGATGGC 735
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 682 GGACCACTGATGATGTGTGTGTATTTGCTTATGCTAGGAAGCTGGAATGATTTCCA 741
OY 736 TCGAAGCTGCT 747
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 742 TCTCTGCTGCT 753
```

RESULT 6

```
US-09-770-445-344
; Sequence 344, Application US/09770445
; Patent No. US20020023281A1
; GENERAL INFORMATION:
; APPLICANT: Gorlach, Jorn
; APPLICANT: An, Yong-Qiang
; APPLICANT: Hamilton, Carol M.
; APPLICANT: Price, Jennifer L.
; APPLICANT: Raines, Tracy M.
; APPLICANT: Yu, Tang
; APPLICANT: Ramezani, Joshua G.
; APPLICANT: Page, Amy
; APPLICANT: Matthew, Abraham V.
; APPLICANT: Ledford, Brooke L.
; APPLICANT: Woessner, Jeffrey P.
; APPLICANT: Haas, William David
; APPLICANT: Garcia, Carlos A.
; APPLICANT: Krieger, Maja
; APPLICANT: Slader, Ted
; APPLICANT: Davis, Keith R.
; APPLICANT: Allen, Keith
; APPLICANT: Hoffman, Neil
; APPLICANT: Hubban, Patrick
; TITLE OF INVENTION: Expressed Sequences of Arabidopsis
; FILE REFERENCE: 202305 (PARA-012PRV)
; CURRENT APPLICATION NUMBER: US/09/770,445
; PRIOR FILING DATE: 2001-01-26
; PRIOR APPLICATION NUMBER: US 60/178,472
; PRIOR FILING DATE: 2000-01-27
; NUMBER OF SEQ ID NOS: 999
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 344
; LENGTH: 948
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-770-445-344
```

Query Match 20.3%; Score 209.2; DB 10; Length 948;
Best Local Similarity 64.9%; Pred. No. 8.5e-48;
Matches 361; Conservative 0; Mismatches 183; Indels 12; Gaps 3;

```
OY 202 TTCACGGGAGAGTCTGTCGTGGGGGAGCAATATTCGACACCGACAGATATCCCGCC 261
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 193 TTCATGAGACTGTCTATGTGTGCGGACACATCTGCACACTGACCAATCATTCGCCG 252
OY 262 GAGCACTCACTCTGTGCTCCCTCCAGCCGAGACGATACCGCAAGCTGCTCTTCCG 321
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 253 GAGTTTCTACTCTGCTCCCTTCGATCCAGAGAAATACGAAATCTCGTTCTTACGT 312
OY 322 TTCGGGGGCTCCATTCGCGGCTTACCGACGCGCTTCTGCTCGGCTGAGAGATCC 381
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 313 TTAGTTGCTTCCA--GCTTTTACAAAGGAGATTCGTTGATGAGCTAGATGAAG 369
OY 382 TCCCGCTACGCAATATTCGCGGAGACCAATTCGAGGTGCGGTTCTCTCGCGAGCAC 441
```

```

Db      | 370  ACGAATACCTCATCATCTATTTGGCCGTGAACAACTTGGATGTGATCTCACGTGAACAT 429
      | 442  GCGCCCGTGGCGCTTGGGGCCGTCGCGACGCCAATTTGSGAGAGGCTACGCCGC 501
      | 430  GCTCCCGTTTGTAGGAGACAGCGGAGCTAAAGCAGTGTGGCTCATCTTTATGATGA 489
      | 502  ATCTTTTTCGCACTCCGTGGCGACTGAGAGAGTGTATCCCTCCAGAGCTACGAGAGT 561
      | 490  ATCTTTTTCAGGAACCTCTTCTACTGTGATGTTTATCTTTGATTT---CTGAAGTT 546
      | 562  GGGGCTCGAAGAGTGAAGACAGCGGAGTGTGTACCGCTTGACCTTGTCTAACTCCG-- 619
      | 547  AGGGTTTGTGATGATGTATCAACCTGTGTATTTGGCGACTGTGTGAGTTGAGGAGAGAT 606
      | 620  ---TTTTATTACCACTCTGTGCGAAGAGTATCAAGCTGAACCAATTGTGTATGCT 675
      | 607  AGTATTTTGTATCATATCATATGAGCTGGAAGAGTATTAAGCTTAAGCCGATTGTGATGCT 666
      | 676  GGGCCTGTATTTAGGCGGAGGAGATCTTTGCTTACGCCCGGAGACAGAGATATTTGCG 735
      | 667  GGAACAGTATGTATGATGCTGTGTATATTTGCTTATGCTAGGAAAGCTGGAATGATTCCA 726
      | 736  TCGAAGCTGTGCTAT 751
      | 727  TCTGTGCTGTGCTGAT 742

```

RESULT 7
US-09-938-842A-1335
; Sequence 1335, Application US/09938842A
; Patent No. US20020160378A1
; GENERAL INFORMATION:

```

; APPLICANT: Harper, Jeff
; APPLICANT: Kreps, Joel
; APPLICANT: Wang, Xun
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
; FILE REFERENCE: SCRIPT300-3
; CURRENT APPLICATION NUMBER: US/09/938,842A
; CURRENT FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: US 60/227,866
; PRIOR FILING DATE: 2000-06-24
; PRIOR APPLICATION NUMBER: US 60/264,647
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/300,111
; PRIOR FILING DATE: 2001-06-22
; NUMBER OF SEQ ID NOS: 5379
; SEQ ID NO 1335
; LENGTH: 771
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
; US-09-938-842A-1335

```

Query Match 19.1%; Score 197.8; DB 9; Length 771;
Best Local Similarity 61.4%; Pred. No. 1.1e-44;
Matches 373; Conservative 0; Mismatches 222; Indels 12; Gaps 3;

```

      | 144  CATGTCCCGCGCGCGCTGTCCGCGGCGGCGAGCACTGCGCGTTCAGCGCGTTT 203
      | 168  CATCCACGTCGTCGCGCTTCAGAAATCCGACTCTAAAGAACCCCTAGCCACACAACTT 227
      | 204  CACGCGGAGTGTGCTGTGCGGAGGACAAATTCAGACCGACGACGATATCCCGCGGA 263
      | 228  CACGCGCTCTGTATGTCTTGAAGAGACAACTATAGACACGACGATCATCCCGCAGG 287
      | 264  GCACTCACTCTGTGCTCTTCAAGCGGAGACGATACCGCAAGCTGCTTCTTGCCTT 323
      | 288  AGCGGCTTGCACCTTCCCATCAACACAGCAGAGCGTATGATGCGCGCTCACGCTCT 347
      | 324  CGGGGGGCTCCATCCGCGGCGCTACCGCGCGCTTGTGCTGCTCGGGGTGAGGTCTCTC 383

```

```

      | 348  CTCTGTCTACCA---GACTTCACAAACACGCTTCAATTGAGCCAGAGACAGATC 404
      | 384  CCGTACGCCATCATTTGTGCGGAGACCCAACTTGGGTGCGTTCTCTCGGAGACGC 443
      | 405  AAGTACTCATATCATATATCCGCGGCGGAAACTTGTGTGGGATCGTCACTCAATATC 464
      | 444  GCGCGTCTGTGGGGCGCTGGCGGACGCGCATTTGTCGAGAGGCTACGCGGCGT 503
      | 465  TCCGGTCTGTCTTGGAGAGCTGTGAGCTAAAGCATATGTTGATGCTTACCAAGAAAT 524
      | 504  CTTTTCGCACTCTCGTGGCCACTGGAAGAGTGTACCTCTGTGAGCTCAGGAGCTTGG 563
      | 525  CTTTTCGCACTCTCGTGGTGTCTACAGGAGGTGTTCCCTCGAG---TCAGAGGTAG 581
      | 564  GGCCTGGAAGAGTGAAGACAGGAGATGTGTACCGGTGACCTTGTCTAACTCCG---- 619
      | 582  AGTCTGTAGAGAGTGTATGACAGAGATATGCGTACATGTAGCTGATGATGTGTG 641
      | 620  ---TTTTATTACCACTCTGTGCGAAGAGTATCAAGCTGAACCAATTGTGTATGCTG 677
      | 642  TTTATGTACTATATACAGACCGGTAATAACTATTAAGCTGAAGTGTGCTGTGCTG 701
      | 678  CCTGTAAATTGAGCGGAGGAGATCTTGCCTACGCCCGGAGACAGAGATGATGTGCTG 737
      | 702  ACCGTTATGTATGCTGTGCTGTGATTTTTCCTTATGCGAGAGATGATGAGATATTCATC 761
      | 738  GAAAGCT 744
      | 762  ATTAGCT 768

```

RESULT 8
US-09-770-445-700

```

; Sequence 700, Application US/09770445
; Patent No. US20020023281A1
; GENERAL INFORMATION:
; APPLICANT: Goriach, Jörn
; APPLICANT: An, Yong-Qiang
; APPLICANT: Hamilton, Carol M.
; APPLICANT: Price, Jennifer L.
; APPLICANT: Raines, Tracy M.
; APPLICANT: Xu, Yang
; APPLICANT: Kameoka, Joshua G.
; APPLICANT: Page, Amy
; APPLICANT: Matthew, Abraham V.
; APPLICANT: Ledford, Brooke L.
; APPLICANT: Moesner, Jeffrey P.
; APPLICANT: Haas, William David
; APPLICANT: Garcia, Carlos A.
; APPLICANT: Krieker, Maja
; APPLICANT: Slader, Ted
; APPLICANT: Davys, Keith R.
; APPLICANT: Allen, Keith
; APPLICANT: Hoffman, Neil
; APPLICANT: Hurban, Patrick
; TITLE OF INVENTION: Expressed Sequences of Arabidopsis
; FILE REFERENCE: 2023US (PARA-012PRV)
; CURRENT APPLICATION NUMBER: US/09/770,445
; PRIOR APPLICATION NUMBER: 2001-01-26
; PRIOR FILING DATE: 2000-01-27
; NUMBER OF SEQ ID NOS: 999
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 700
; LENGTH: 829
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
; US-09-770-445-700

```

Query Match 19.0%; Score 196.6; DB 10; Length 829;
Best Local Similarity 60.3%; Pred. No. 2.4e-44;
Matches 382; Conservative 1; Mismatches 235; Indels 15; Gaps 3;

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 3, 2003, 02:53:39 ; Search time 1498 Seconds

(without alignments)
1168.183 Million cell updates/sec

Title: US-10-027-450-46

Perfect score: 1033

Sequence: 1 atcatgagcgagctctgtc.....aaaaaaaaaaaaaaaa 1033

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

```
EST:*
1: em_estha:*
2: em_esthm:*
3: em_estin:*
4: em_estom:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estom:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rod:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	910.4	88.1	1085	11 AY109732	AY109732 zea mays
2	570.6	55.2	571	9 AY1977840	AY1977840 496028H06
3	536.8	51.9	546	9 AY1820332	AY1820332 605091F07
4	533.8	51.7	589	9 AY1833533	AY1833533 605091F07
5	526.6	51.0	556	10 BE510562	BE510562 946053G08
6	493	47.7	568	10 BE510561	BE510561 946053G08

C	7	472	45.7	601	9	AT677354	AT677354 605053H06
	8	471.6	45.7	534	10	BE597022	BE597022 P11_60_DO
	9	461.2	44.6	564	10	BE360232	BE360232 DGI_62_GO
	10	453.4	43.9	566	10	AM677919	AM677919 WS1_12_DO
	11	453	43.9	551	12	AM677919	AM677919 WS1_12_DO
	12	442.6	42.8	546	10	BE360264	BE360264 DGI_62_BO
	13	435.4	42.1	841	12	BE309960	BE309960 HVSMEC001
	14	428.8	41.5	534	10	AM747153	AM747153 WS1_66_GO
	15	412.6	39.9	500	12	BE267153	BE267153 1000112BO
	16	410.6	39.7	499	10	BE051642	BE051642 za86f09.b
	17	409	39.6	409	13	BE1096798	BE1096798 949015C02
	18	403.6	39.1	615	10	BE360295	BE360295 DGI_62_GO
	19	388.8	37.6	691	14	BO838113	BO838113 WHE2906.F
	20	387.2	37.5	451	10	BE051643	BE051643 za86f09.g
	21	376.6	36.5	403	9	AT438914	AT438914 486014G05
	22	371.8	36.0	700	9	AL508419	AL508419 AL508419
	23	371.4	36.0	393	13	BM499030	BM499030 949015C02
	24	366	35.4	635	10	AM678025	AM678025 WS1_12_DO
	25	365.4	35.4	619	10	BE596939	BE596939 P11_60_DO
	26	335	32.4	546	9	AT964511	AT964511 496012D10
	27	330.8	32.0	557	12	BE048147	BE048147 OVI_28_CO
	28	326.4	31.6	866	12	BE621774	BE621774 HVSMEC001
	29	323.4	31.3	700	9	AL508228	AL508228 AL508228
	30	322.2	31.3	710	10	BE54517	BE54517 HVSMEC009
	31	310.8	30.1	373	9	AT902008	AT902008 618013F09
	32	300.4	29.1	615	12	BE619744	BE619744 HVSMEC000
	33	289.8	28.1	580	14	BM732906	BM732906 sa191b04.y
	34	284.2	27.5	565	12	BE804009	BE804009 sr181h2.y
	35	283.4	27.4	369	12	BE656248	BE656248 FM1_44_HO
	36	269	26.0	479	12	BE478573	BE478573 WHE2003.B
	37	258.8	25.1	834	14	BO797019	BO797019 EST_5957
	38	254.4	24.6	286	17	BH221062	BH221062 1006099B0
	39	244.4	23.7	714	9	AT096006	AT096006 AT096006
	40	242	23.4	640	13	BE353523	BE353523 HVSMEC001
	41	231.6	22.4	596	10	AM649879	AM649879 EST328333
	42	228	22.1	717	17	BH883347	BH883347 h43h08.b
	43	228	22.1	746	17	BE695275	BE695275 BOH7677R
	44	226.6	21.9	1344	10	BE422061	BE422061 HMM019CE
	45	226.4	21.9	524	13	BE498836	BE498836 sa123b12.

ALIGNMENTS

RESULT 1
LOCUS AY109732 1085 bp mRNA linear HTC 25-MAY-2002
DEFINITION Zea mays CL1042_1 mRNA sequence.
ACCESSION AY109732
VERSION AY109732.1 GI:21213564
KEYWORDS HTC.
SOURCE Zea mays.
ORGANISM Zea mays

REFERENCE
AUTHORS Hainey,C.F., Doan,M., Miao,G.H., Vogel,J.M., Whitsitt,M.S.,
Arthur,L.W., Hanafey,M., Morigane,M. and Tingey,S.V.
TITLE Maize Mapping Project/Dupont Consensus Sequences for Design of
1 (bases 1 to 1085)

JOURNAL
REFERENCE 2 (bases 1 to 1085)
AUTHORS Direct Submission
TITLE Submitted (25-APR-2002) Maize Mapping Project, University of
JOURNAL Missouri, Columbia, MO 65211, USA

FEATURES
SOURCE location/Qualifiers
1..1085
/organism="Zea mays"
/db_xref="Waizdb:629913"
/db_xref="taxon:4577"
/clone="CL1042_1"

/clone.lib="Maize Mapping Project/DuPont Consensus Library"
 /note="this sequence is part of a project of EST assemblies resulting from the application of public configs to seed Dupont configs; this resource was assembled by Dupont as part of a collaboration for the overgo addressing of BACs in conjunction with the Maize Mapping Project"

BASE COUNT 203 a 303 c 295 g 239 t 45 others
 ORIGIN

Query Match 88.1%; Score 910.4; DB 11; Length 1085;
 Best Local Similarity 94.0%; Pred. No. 1.4e-136;
 Matches 941; Conservative 1; Mismatches 52; Indels 7; Gaps 2;

1 ATCATGGCGCGGCTCTGTCTGCGGGAGCGGGGTCTCCAGCGAGCGCTTCTAGACCCCAATC 60
 |||||
 8 ATCATGGCGCGGCTCTGTCTGCGGGAGCGGGGTCTCCAGCGAGCGCTTCTAGACCCCAATC 67
 |||||
 61 CGAGCTCCAAACGAGCGCGTTTATCCGGGCTCCAGCTCACTGTTCATCGGCTCCACATCA 120
 |||||
 68 CGAGCTCCAAACGAGCGCGTTTATCCGGGCTCCAGCTCACTGTTCATCGGCTCCACATCA 127
 |||||
 121 CTAAATGCGCGCGCGCTGGGTCTCATCTCCCGCGCGCTGTCTGCCGCGGCGGACAC 180
 |||||
 128 CTAAATGCGCGCGCGCTGGGTCTCATCTCCCGCGCGCTGTCTGCCGCGGCGGACAC 187
 |||||
 181 AGCTCGCGGTGTCTGAGCGGTTTTCACAGCGGAGTCTGTGTGTGGGCGACAATTCAC 240
 |||||
 188 AGCTCGCGGTGTCTGAGCGGTTTTCACAGCGGAGTCTGTGTGTGGGCGACAATTCAC 247
 |||||
 241 ACCGACCAAGATCCCGCGCGAGACACTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 300
 |||||
 248 ACCGACCAAGATCCCGCGCGAGACACTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 307
 |||||
 301 CGCAAGCTCGGTTCTTCTGAGCTTCTGCGGGGCTCCCATCCGCGGCTTACCAGCGCGTTC 360
 |||||
 308 CGCAAGCTCGGTTCTTCTGAGCTTCTGCGGGGCTCCCATCCGCGGCTTACCAGCGCGTTC 367
 |||||
 361 GTCCGTCGGGTGTGAGGAGTCTCCCGCTACGCGCATATGTGGGGGAGCAACATTCCGG 420
 |||||
 368 GTCCGTCGGGTGTGAGGAGTCTCCCGCTACGCGCATATGTGGGGGAGCAACATTCCGG 427
 |||||
 421 TGGGTTCTCTCTCGAGACAGCGCGCTTGGGCGGCTTGGGCGACGCGGCATT 480
 |||||
 428 TGGGTTCTCTCTCGAGACAGCGCGCTTGGGCGGCTTGGGCGACGCGGCATT 487
 |||||
 481 GTTGGGAGGCTTACGCGCGCATCTTTTTCGCAACTCCGTGGCCACTGGAAGGTGTAC 540
 |||||
 488 GTTGGGAGGCTTACGCGCGCATCTTTTTCGCAACTCCGTGGCCACTGGAAGGTGTAC 547
 |||||
 541 CCTCGGAGCTCACGAGAGTGGGGCTTGAAGAGTGAAGAGGAGTGTGTAC 600
 |||||
 548 CCTCGGAGCTCACGAGAGTGGGGCTTGAAGAGTGAAGAGGAGTGTGTAC 607
 |||||
 601 GTGACCTTGTCTAACTCCGTTTATTAACCACTCTCTGGCAAGAGTGAAGAGTGAAG 660
 |||||
 608 GTGACCTTGTCTAACTCCGTTTATTAACCACTCTCTGGCAAGAGTGAAGAGTGAAG 667
 |||||
 661 CCAATGTGTGAGTGGGCGCTGTAAATGAGGCGGAGGAGTCTTGTGCTACGCGCGGAAG 720
 |||||
 668 CCAATGTGTGAGTGGGCGCTGTAAATGAGGCGGAGGAGTCTTGTGCTACGCGCGGAAG 727
 |||||
 721 ACAGAGATGATGCGTCCAAAGCTGTGCATGAGGGAA---AGCTTATGACGCGGAGC 775
 |||||
 728 ACAGAGATGATGCGTCCAAAGCTGTGCATGAGGGAAAGATGAGCTTAAAGGAGGAGC 787
 |||||
 776 CTCTCGGAGATGAAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 835
 |||||
 788 CTCTCGGAGATGAAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 847
 |||||
 836 GTGACGAGTGTCTCAAAATGATGCGGCTTACGGAATATGATGAATCAATCAATTG 895
 |||||

Db 848 GTGACGAGTGTCTCAAAATGATGCGGCTTACGGAATATGATGAATCAATCAATTG 907
 |||||
 QY 896 GTCTTGTTCACAGATCGTTTTTTTGTACTAGTA--CTGTCAATTTACTCTGCC 953
 |||||
 Db 908 GTCTTGTTCACAGATCGTTTTTTTGTACTAGTA--CTGTCAATTTACTCTGCC 967
 |||||
 QY 954 TGTCTACTGTCTTATCTGTGTAATTAAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 994
 |||||
 Db 968 TGTCTACTGTCTTATCTGTGTAATTAAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1008
 |||||

RESULT 2
 AI977840/c 571 bp mRNA linear EST 27-AUG-1999
 LOCUS 496028406.xl 496 - stressed shoot cDNA library from Wang/Bohnert
 DEFINITION lab zea mays cDNA, mRNA sequence.
 ACCESSION AI977840
 VERSION AI977840.1 GI:5791048
 KEYWORDS EST.
 SOURCE Zea mays.
 ORGANISM Zea mays.
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
 1 (bases 1 to 571)

REFERENCE
 AUTHORS Maize ESTs from various cDNA libraries sequenced at Stanford
 TITLE University
 JOURNAL Unpublished (1999)
 COMMENT Contact: Walbot V
 Department of Biological Sciences
 Stanford University
 855 California Ave, Palo Alto, CA 94304, USA
 Tel: 650 723 2227
 Fax: 650 725 8221
 Email: walbot@stanford.edu
 Plate: 496028 row: H column: 06.
 Location/Qualifiers

FEATURES
 source
 1..571
 /organism="Zea mays"
 /cultivar="B73"
 /db_xref="taxon:4577"
 /clone.lib="496 - stressed shoot cDNA library from
 Wang/Bohnert lab"
 /issue-type="seedling"
 /dev-stage="salt stress"
 /lab_host="E.coli XL Gold"
 /note="Organ: shoot; Vector: pBluescriptII SK(+); XR;
 Wang/Bohnert"

BASE COUNT 150 a 162 c 135 g 124 t
 ORIGIN

Query Match 55.2%; Score 570.6; DB 9; Length 571;
 Best Local Similarity 99.8%; Pred. No. 4.4e-82;
 Matches 570; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 411 CAATCGGGGTGGGTTCTCTCGGAGCAGCGCGCGCTTGGGCGCGTGGGCGC 470
 |||||
 Db 571 CAATCGGGGTGGGTTCTCTCGGAGCAGCGCGCGCTTGGGCGCGTGGGCGC 512
 |||||
 QY 471 ACGGCGCATTTGTGCGSAGGAGTACGCGCGCATCTTTTTCGCACTCCGTGGCCACTG 530
 |||||
 Db 511 ACGGCGCATTTGTGCGSAGGAGTACGCGCGCATCTTTTTCGCACTCCGTGGCCACTG 452
 |||||
 QY 531 AGAGGTGACCTTGTGAGAGTCAAGAGTGTGGGGCGCTGGAAGAGTGAAGAGGAGGA 590
 |||||
 Db 451 AGAGGTGACCTTGTGAGAGTCAAGAGTGTGGGGCGCTGGAAGAGTGAAGAGGAGGA 392
 |||||
 QY 591 TGTGTGACGCTGAGACCTTGTCTAACTCGTTTTTATTAACCACTCTGGAAGAGTA 650
 |||||
 Db 391 TGTGTGACGCTGAGACCTTGTCTAACTCGTTTTTATTAACCACTCTGGAAGAGTA 332
 |||||
 QY 651 CAAGCTGAACCAATTTGTATGCTGGCCTGTAAATGAGGCGGAGGAGTCTTCCCTA 710
 |||||

[illegible][illegible]

REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
1 (bases 1 to 546)	Walbot V.	Maize ESTs from various cDNA libraries sequenced at Stanford	University	Unpublished (1999)
			Contact: Walbot V	Department of Biological Sciences
				Stanford University
				855 California Ave, Palo Alto, CA 94304, USA
				Tel: 650 723 2227
				Fax: 650 725 8221
				Email: walbot@stanford.edu
				Place: 605091 row: F column: 07.
FEATURES				Location/Qualifiers

```

/organism="Zea mays"
/cultivar="Ohio43"
/db_xref="taxon:4577"
/db_clone_lib="605 - Endosperm cDNA library from Schmidt lab"
/tissue_type="uncellular, embryo, and endosperm"
/dev_stage="10-14 days post-pollination"
/lab_host="DH5 (alpha)"
/notes="Organ: Kernel; Vector: pMD-GM4-2'; Site_1: ECORI;
Site_2: XhoI; Kernel endosperm cDNA library from Schmidt
lab"

```

Query Match	51.9%	Score 536.6	DB 9	Length 546
Best Local Similarity	98.7%	Pred. No 1.2e-76		
Matches 539	Conservative 1	Mismatches 6	Indels 0	Gaps 0
QY	67	CCACACGAGCGCTTATCGGCGGCTCCACGCTCACCTGTATGTCGCTCCACTCATCTAAAA	126	
Db	1	CCAACGAGGCGGCTTATCGGCGGCTCCACGCTCATGTCATGCTCCACTCATCTAAAAA	60	

QY	127	TGCGCGCGGCTTGGATCATCTTCTCCCGCGGCGCTGCTGCGGGGGGACACACTGTG	186
Db	61	TGCGCGCGGCGCTTGGGCCCATCTCTCCCGCGGCGCTGCTGCGGGGGGACACACTGTG	120
QY	187	CCGTCTGACGCGGTTTTTCCAGGGGAGTGTCTTGTTGGGCGCAATATGACACCGAC	246
Db	121	CCGTCTGACGCGGTTTTTCCAGGGGAGCGTGTGGTTGGGCGCAATATGACACCGAC	180
QY	247	CAGATCATCCCGCGGAGCACTCATCTGTGGTGGCTTCGAAGGCGGGAGTGATCCGCAAG	306
Db	181	CAGATCATCCCGCGGAGCACTCATCTGTGGTGGCTTCGAAGCGGGAGTGATCCGCAAG	240
QY	307	CTCGGTTCTTGGCTTTCGGGGGGCTCCCATCGCGGCGCTACCGGAGCGCGTTTGTGTGT	366
Db	241	CTCGGTTCTTGGCTTTCGGGGGGCTCCCATCGCGGCGCTACCGGAGCGCGTGTGTGT	300
QY	367	CCGGGTTGAGGAGTCCCTCCCGGTAGCGCATCATTTTTCGGGAGGAGGCAATCTTCGGGTGGGT	426
Db	301	CCGGGTTGAGGAGTCCCTCCCGGTAGCGCATCATTTGTTCGGGAGGAGCAATCTTCGGGTGGGT	360
QY	427	TTCCTCTTCGCGAGCACGCGCCGTGCGCGTTTGGGGCGGCTGGAGCGACGCGCATTTGTGTGCS	486
Db	361	TTCCTCTTCGCGAGCACGCGCCGTGCGCGTTTGGGGCGGCTGGAGCGACGCGCATTTGTGTGTG	420
QY	487	GAGGGCTAGGCGCGGATCTTTTGTGGCAATCTGTGGGCGACTGGAGAGGGTATCCCTGTG	546
Db	421	GAGGGCTAGGCGCGGATCTTTTGTGGCAATCTGTGGGCGACTGGAGAGGGTATCCCTGTG	480
QY	547	GAGTCACGAGACTTGGGGGCTGGAAGGAGTGCAAGACAGGGGATGTGGTCACTCGTGAG	606
Db	481	GAGTCACGAGACTTGGGGGCTGGAAGGAGTGCAAGACAGGGGATGTGGTCACTCGTGAG	540
QY	607	CTTGGCT 612	
Db	541	CTTGGCT 546	

RESULT	4
AIR8353/c	
LOCUS	AIR83533
DEFINITION	605091F07.x2 60S - EndospERM CDNA library from Schmidt lab Zea mays.
ACCESSION	AIR83533
VERSION	AIR83533
KEYWORDS	EST. GI:5455843
SOURCE	Zea mays.
ORGANISM	Zea mays

REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
1 (bases 1 to 589)	Malbot V.	Maize ESTs from various cDNA libraries sequenced at Stanford	University Unpublished (1999)	
		Contact: Malbot V		
		Department of Biological Sciences		
		Stanford University		
		855 California Ave, Palo Alto, CA 94304, USA		
		Tel: 650 723 2227		
		Fax: 650 725 8221		
		Email: walpote@stanford.edu		
		Plate: 605091 row: F column: 07.		
FEATURES		Location/Qualifiers		
source		1..589		

```

/clone_lib="605 - Endosperm cDNA library from Schmidt lab"
/rissue_type="nucellar embryo, and endosperm"
/dev_stage="10-14 days post pollination"
/lab_host="DH5(alpha)"
/note="Organ: Kernel; Vector: pMD-GAL4-2'; Site_1: EcoRI;

```


VERSION BE510561.1 GI:9731809
KEYWORDS EST.
SOURCE Zea mays.
ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoideae; Andropogoneae; Zea.
REFERENCE 1 (bases 1 to 568)
AUTHORS Walbot,V.
TITLE Maize ESTs from various cDNA libraries sequenced at Stanford University
JOURNAL Unpublished (1999)
COMMENT Contact: Walbot V
Department of Biological Sciences
Stanford University
855 California Ave, Palo Alto, CA 94304, USA
Tel: 650 723 2227
Fax: 650 725 8221
Email: walbot@stanford.edu
Plate: 946053 row: G column: 08.
Location/Qualifiers
1. 568
/organism="Zea mays"
/cultivar="OH43"
/db_xref="taxon:4577"
/clone_lib="946 - tassels primordialium prepared by Schmidt lab"
/tissue_type="tassels"
/dev_stage="just after the transition from vegetative to inflorescence development"
/lab_host="XLOLR"
/note="Organ: tassels; Vector: HybridAP; Site.1: EcoRI; Site.2: XhoI; George Chuck dissected immature tassels between 1mm and 3mm. Sharon Stafield prepared the cDNA library in HybridAP. Sample insert size range was 350 bp to 3 kb with a 1 kb average."

BASE COUNT 155 a 154 c 129 g 130 t
ORIGIN

Query Match 47.7%; Score 493; DB 10; Length 568;
Best Local Similarity 97.4%; Pred. No. 1.1e-69;
Matches 523; Conservative 1; Mismatches 6; Indels 7; Gaps 2;

OY 465 TGGCGAGCGCCATTGTTGSGAGGGCTACGCGGCACTTTTTCGCAACCTCGTGGC 524
DB 568 TGGCGAGCGCCATTGTTGSGAGGGCTACGCGGCACTTTTTCGCAACCTCGTGGC 509
OY 525 CACTGAGAGGTGTACCCCTGAGCTCAGGAGCTTGGGCGTGGAGAGTGCAGAC 584
DB 508 CACTGAGAGGTGTACCCCTGAGCTCAGGAGCTTGGGCGTGGAGAGTGCAGAC 449
OY 585 AGGGATGTGTACCGCTGAGCTTGTCTAATCGTTTATTACCAACACCTCTGGCAA 644
DB 448 AGGGATGTGTACCGCTGAGCTTGTCTAATCGTTTATTACCAACACCTCTGGCAA 389
OY 645 GGAGTCAAGCTGAAGCAATGTGTGATGCTGCGCTTAATGAGGGGAGGATCTT 704
DB 388 GGAGTCAAGCTGAAGCAATGTGTGATGCTGCGCTTAATGAGGGGAGGATCTT 329
OY 705 TGGCTACGCCCGGAGACAGATGATGCTGGAAGCTGCTGATGAGGAA-----A 759
DB 328 TGGCTACGCCCGGAGACAGATGATGCTGGAAGCTGCTGATGAGGAAAGATCA 269
OY 760 GCTTATGACGCGGAGCTTGGGAGATGAAGAAATAGCTGAGATTAAGATTAAGATT 819
DB 268 GCTTATGACGCGGAGCTTGGGAGATGAAGAAATAGCTGAGATTAAGATTAAGATT 209
OY 820 ACTGACACTACTGATGTGAGAGGTGTCTCAAAATAGTTGCGGCTACCGAAATTTAGA 879
DB 208 ACTGACACTACTGATGTGAGAGGTGTCTCAAAATAGTTGCGGCTACCGAAATTTAGA 149
OY 880 TGAATCAATCAATTGGTCTTTGTTCACAGATGTTTTTTT--TTTGTACTGACTGTGA 937
DB 148 TGAATCAATCAATTGGTCTTTGTTCACAGATGTTTTTTTGTTCAGTACTGTGA 89

OY 938 CAATTACTCTGCGCTGCTACTGTTTATATGTTGCAATGACTGCTGTGGCA 994
DB 88 CAATTACTCTGCGCTGCTACTGTTTATCTGTTGCAATGACTGCTGTGGCA 32

RESULT 7
AI677354/c
LOCUS AI677354 601 bp mRNA linear EST 02-FEB-2000
DEFINITION 605053H06.x1 605 - Endosperm cDNA library from Schmidt lab Zea mays
ACCESSION AI677354
VERSION AI677354.1 GI:4886234
KEYWORDS EST.
SOURCE Zea mays.
ORGANISM Zea mays.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoideae; Andropogoneae; Zea.
REFERENCE 1 (bases 1 to 601)
AUTHORS Walbot,V.
TITLE Maize ESTs from various cDNA libraries sequenced at Stanford University
JOURNAL Unpublished (1999)
COMMENT Contact: Walbot V
Department of Biological Sciences
Stanford University
855 California Ave, Palo Alto, CA 94304, USA
Tel: 650 723 2227
Fax: 650 725 8221
Email: walbot@stanford.edu
Plate: 605053 row: H column: 06.
Location/Qualifiers
1. 601
/organism="Zea mays"
/cultivar="Ohio43"
/db_xref="taxon:4577"
/clone_lib="605 - Endosperm cDNA library from Schmidt lab"
/tissue_type="nucellar, embryo, and endosperm"
/dev_stage="10-14 days post-pollination"
/lab_host="DH5(alpha)"
/note="Organ: kernel; Vector: PAD-GAL4-2'; Site.1: EcoRI; Site.2: XhoI; Kernel endosperm cDNA library from Schmidt lab"

BASE COUNT 172 a 161 c 129 g 139 t
ORIGIN

Query Match 45.7%; Score 472; DB 9; Length 601;
Best Local Similarity 96.9%; Pred. No. 2.6e-66;
Matches 504; Conservative 1; Mismatches 6; Indels 9; Gaps 2;

OY 484 GCGAGCGCTACGCGGCACTTTTTCGCAACCTCGGCGCACTGGAGAGGTACCT 543
DB 601 GCGAGCGCTACGCGGCACTTTTTCGCAACCTCGGCGCACTGGAGAGGTACCT 542
OY 544 CTGAGCTCAGGAGCTTGGGCTGGAAGAGTGCAGAGAGGAGTGTGTACCGTG 603
DB 541 CTGAGCTCAGGAGCTTGGGCTGGAAGAGTGCAGAGAGGAGTGTGTACCGTG 482
OY 604 GACCTTGCTAATCGCTTTTATTAAACACACCTCTGGCAAGAGTCAAGCTGAACCA 663
DB 481 GACCTTGCTAATCGCTTTTATTAAACACACCTCTGGCAAGAGTCAAGCTGAACCA 422
OY 664 ATTGTGATGTGGCCCTGTAATTGAGGCGGAGGAGATCTTGGCTACGCCCGGAACA 723
DB 421 ATTGTGATGTGGCCCTGTAATTGAGGCGGAGGAGATCTTGGCTACGCCCGGAACA 362
OY 724 GGAATGATGGTGAAGAGCTGTCATGAGAGGAA-----AGCTTATGACGCGAGGCTC 778
DB 361 GGAATGATGGTGAAGAGCTGTCATGAGAGGAAAGATCAAGCTTATGACGCGAGGCTC 302
OY 779 TCGGAGATGAAGAGTAAGCTGAGATTAGACTTAAGATTACTGACCACTTGTATGTC 838

	DB	301	TGCGAATGAAGAAGTAACCTGCAGTATTAGAGACTAAAGATTACGTACCACTTGATGTC	242
	OY	839	GAGCGTGTCCAAAATAAAGTTGGCGCTCCGCCAAATTTTGATGAAATCAATCAATTGGCTC	898
	DB	241	GACGGTGTCTCAAATAATAGTTGGCGCTCCGCCAAATTTGATGATCAATCAATTTGGCT	182
	OY	899	TTTGTCACAGATGGTTTTT---TTTGTACTAGTACTTGTACAATTTGACTCTGGCC	954
	DB	181	TTTGTACACATGGTTTTTTTTTTTGTGTACTACTGTTACAATTTGACCTCGGCT	122
	OY	955	GCTACTGTTCTATCTGTTTGAAATACCTCTGTGGCA	994
	DB	121	GCTACTGTTCTATCTGTTTGAAATACCTCTGTGGCA	82
	RESULT 8			
	BE597022		534 bp	mRNA
	LOCUS		Pil.60.D05_b1_A002 Pathogen induced 1 (Pil)	Sorghum bicolor cDNA,
	DEFINITION			
	ACCESSION		BE597022	
	VERSION		BE597022.1	GI:9852095
	KEYWORDS		EST.	
	SOURCE		sorghum.	
	ORGANISM		Sorghum bicolor	
	REFERENCE		Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;	
	AUTHORS		Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC	
			clade; Panicoideae; Andropogoneae; Sorghum.	
			1 (bases 1 to 534)	
			Cordonnier-Pratt,M.-M., Gingle,A., Dean,R., Sudman,W. and Pratt	
			,L.H.	
	TITLE		An EST database for Sorghum: pathogen-induced plants	
	JOURNAL		Unpublished (2000)	
	COMMENT		Contact: Cordonnier-Pratt MM	
			Laboratory for Genomics and Bioinformatics	
			The University of Georgia, Department of Plant Biology	
			Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA	
			Tel.: 706 542 1860	
			Fax: 706 583 0210	
			Email: mmp@prattuga.edu	
			Sequences have been trimmed to exclude PolyA, vector and regions	
			below phred quality 16. The threshold for highest quality sequence	
			is 20.	
			Seq primer: JEN REV	
			High quality sequence stop: 533	
			POLYA-No.	
	FEATURES			
	SOURCE			
			Location/Qualifiers	
			1..534	
			/organism="Sorghum bicolor"	
			/db_xref="taxon:4558"	
			/clone_lib="Pathogen induced 1 (PIL)"	
			/note="Organ: Anthracnose-infected leaves from	
			two-week-old sorghum plants 48 hr after inoculation;	
			vector: pBluescript II from lambda Zap II; Site.1: XhoI;	
			Site.2: EcoRI; two-week-old sorghum plants (BRX 623	
			cultivar) were infected with pathogen (isolate PMW42I of	
			Colletotrichum graminicola, which is a sorghum isolate).	
			RNA was prepared from infected leaves harvested from 45	
			seedlings (2 weeks old) exhibit juvenile resistant	
			reaction, which is an incompatible interaction. As they	
			grow older (4 weeks or older), plants resume susceptibility	
			to anthracnose disease. The library was made from poly-A	
			RNA in the cloning vector lambda Zap II. Clones to be	
			sequenced were prepared by mass excision. WARNING: While	
			most or all ESTs are expected to derive from the host	
			plant, no effort was made to eliminate ESTs deriving from	
			the pathogen."	
	BASE COUNT		95 a	173 c 167 g 99 t
	ORIGIN			
	Query Match		45.7%; Score 471.6; DB 10; Length 534;	

FEATURES	source
Best Local Similarity	93.2%; Pred. No. 3,1e-66; Mismatches 492; Conservative 1; Mismatches 35; Indels 0; Gaps 0;
Db	226 GGGCAACAAATTCACACACCGACGATCACTCCCGCGGAGCAACCTACCTGATGTCGCCCTCC 285
Db	7 GGGGACAAACATCGACACCGACCGACGATCACTCCCGCGGAGCAACCTACCTGATGTCGCCCTCC 66
QY	286 AAGCGGAGCAGATACCGCAAGCTCGGTTCTCTTCCGCTTCCGCGGGGACCTCCATCCCGGGCC 345
Db	67 AAGCTGACAGATACCGCAAGCTCGGCTCTCTTGGCTTCCGCGGGGACCTCCCTCCCGGGCC 126
QY	346 TACCCGAGGCGCTTCGCTCCGGGTAGAGATCTCCCGCTACCGCATCATTTGTCGGC 405
Db	127 TACCCGAGGCGCTTCGCTCCGGGTAGAGATCTCCCGCTACCGCATCATTTGTCGGT 186
QY	406 GGAGCCCAACTTGGGATGCGGTTCTCTTCGAGCAAGCGCCCGTTCGCGCTTGGGGCCGCT 465
Db	187 GGGGCAACAACTTGGGATGCGGTTCTCTTCGAGCAAGCGCCCGTTCGCGCTTGGGGCCGCT 246
QY	466 GGGCGACGCGCAATTTGTTGSGAGGGCTACGGCGCATCTTTTTCGCACTCCGTGGCC 525
Db	247 GGAGCCCGCGCGCTCTCGCGGAGGGCTACGGCGCATCTTTTTCGCACTCCGTGGCC 306
QY	526 ACTGGAGAGGTGTACCCCTCTGTGAGCTCACGGACGTTGGGGCCCTGGAAGAGATGCAAGCA 585
Db	307 ACTGGAAGAGGTGTACCCCTCTGTGAGCTCACGGACGTTGGGGCCCTGGAAGAGATGCAAGCA 366
QY	586 GGGGATTTGGTCACCCGCGGACCTTGGTATACCTCCGTTTTATTTAACACACCTCTGGCAAG 645
Db	367 GGGGATTTGGTCACTCCGCGGACCTTGAATTAATCTCCGTTTTATTTAACACACCTCTGGCAAG 426
QY	646 GAGTACAGCTGAACCAATTTGGTGTGCTGGCCCTGTAAATTGAGGGCGGAGGAGATCTTT 705
Db	427 GAGTACAGCTGAACCAATTTGGTGTGCTGGCCGCGGATTTGAGGGCGGAGGAGATCTTT 486
QY	706 GCTTACGCCCGGAAGACAGAAATGATTTGGTGTGAAGAGCTGCGCATGA 753
Db	487 GCTTATCTCTGGAAGACAGAAATGATTTGGTGTGAAGAGCTGCGCATGA 534
RESULT 9	BE360232 564 bp mRNA linear EST 20-JUL-2000
LOCUS	BE360232
DEFINITION	D01.62.G01.b1_A002 Dark Grown 1 (DG1) Sorghum bicolor cDNA, mRNA
ACCESSION	BE360232
VERSION	BE360232.1
KEYWORDS	EST.
SOURCE	Sorghum.
ORGANISM	Sorghum bicolor
REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoideae; Andropogoneae; Sorghum.
AUTHORS	1 (Phases 1 to 564) Cordonnier-Pratt,M.-W., Gingle,A., Marala,C., Sudman,M. and Pratt ,L.H.
TITLE	An EST database from Sorghum: dark-grown seedlings
JOURNAL	Unpublished (2000)
COMMENT	Contact: Cordonnier-Pratt MM Laboratory for Genomics and Bioinformatics The University of Georgia, Department of Plant Biology Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA Tel: 706 542 1860 Fax: 706 583 0210 Email: mmp@prattuga.edu
FEATURES	Seq primer: JEM REV High quality sequence stop: 543 POLYA-No.
source	Location/Qualifiers 1..564

```

/organism="Sorghum bicolor"
/db_xref="taxon:4558"
/clone_1fb="Dark Grown 1 (DGL)"
/note="Organ: 5-day-old dark-grown seedlings; Vector:
Lambda Zap; Site_1: XhoI; Site_2: EcoRI; The library was
made from poly-A RNA in the cloning vector lambda Zap II
Clones to be sequenced were prepared by mass excision."
BASE COUNT
ORIGIN
73 a 223 c 170 g 98 t

```

Query Match	44.68;	Score 461.2;	DB 10;	Length 564;
Best Local Similarity	92.08;	Pred. No. 1.4e-64;		
Matches 497; Conservative	1;	Mismatches 39;	Indels 3;	Gaps 1.

QY	1	ATCATGGCGGGGCGCTCTGTGCGGGAGCGCGGTGTCCACGGCAGCGCTTTCTAAGCCCCATTC	60
Db	25	ATCATGGCGGGCTCTCGGTGGCGAGCGGGGGTGTCCACGGGGGGCTTTAAGCCCCATTC	84
QY	61	CGAGCTTCCAAACGAGCGCGTTTATCCGGGGCTCCGAGCTCACCTGTCAATGCGCTTCACTCA	120
Db	85	CGAGCTTCCAAACGAGCGGTTCCTCGGGGGCACCGGTCAACCGTCAACCGGAGAACATCC	144
QY	121	CTTAAATAGCGCGCGCGCTGGGTTCATCGTCCCCGGCGCGCTGCTGCCCGCGCGG--GC	177
Db	145	CTTAAATAGCGCGCGCGCTGCTCCATCGTCCCCGGCGCGCTGTGCGCGCGCGCGCGAGC	204
QY	178	AGCAGCTCGCGCTGTCTCAGCGCTTTTTCACGGCGAGTCTTCTGTGTGGCGGACAAATATC	237
Db	205	AGCAGCTCGCGCTGTCTGCTGCCGTTTTCACGGCGAGTCTTGTGTGTGGGAGAACATTC	284
QY	238	GACACCGACCAATATATCCCCCGCAGACACTCATCTGTGTGGCCCTCCAGCGGAGCAG	297
Db	265	GACACCGACCAATATATCCCCCGGAGGACACTCATCTGTGTGGCCCTCCAGACCTGAGACG	324
QY	298	TACCGCAGAGTCAGTCTCTTCGCTTCGCGGGGGCTCCCATCGCGGGCTTACCGGAGCG	357
Db	325	TACCGCAGAGTCAGTCTTTCGCTTCGCGGGGGCTCCCTCGCGGGCTTACCGGAGCGG	384
QY	358	TTTCGCTCGCGGTGAGGAGTCTTCGCGCTACGCCATCATTTGTGGCGGAGCCCAATTC	417
Db	385	TTTCGCTCGCGGGTGAAGGAGTCAACCCGCTACGCCGTCGTCTGTGGGGGCAATTC	444
QY	418	GGGAGCGGTTCTCTGTGCGAGACAGCGCGCGTGGGGTGGGGCGCGGAGCGGCGC	477
Db	445	GGGAGCGGCTCTCTCGCGAGACAGCGCGCGTGGGGTGGGGCGCGGAGCGGCGGCGC	504
QY	478	ATTGTTGCSAGGGCTACGCGCGCATCTTTTTCGAACTCCGTGGCCATCTGAGAGGTG	537
Db	505	GTCGTCGGGAGGGCTACGCGCGCATCTTTTTCGCAACTCCGTGGGCGACTGAGAGAGGTG	564

[illegible]

Tel: 706 542 1860
Fax: 706 583 0210
Email: impratt@uga.edu
Sequences have been trimmed to exclude POLYA, vector and regions below Phred quality 16. The threshold for highest quality sequence is 20.
Seq primer: JEN REV
High quality sequence stop: 565
POLYA=No.

FEATURES	Location/Qualifiers
source	1. .566

BASE COUNT	93	a	196	c	165	g	112	t
ORIGIN	/organism="Sorghum bicolor" /db_xref="taxon:4558" /clone_lib="Water-stressed 1 (WS1)" /note="Organ: Mix of 5-week old plants on days 7 & 8 after water was withheld; Vector: Lambda Zap; Site.1: XhoI; Site.2: EcoRI; The library was made from poly-A RNA in the cloning vector lambda Zap II. Clones to be sequenced were prepared by mass excision."							

Query Match	43.9%;	Score 453.4;	DB 10;	Length 566;
Best Local Similarity	93.5%;	Pred. No. 2.5e-63;		
Matches 472; Conservative	1;	Mismatches 32;	Indels 0;	Gaps 0;

QY	176	GCACACAGCTGCAGCTGCAGCCGCTTTTCCACGGCAGTAGCTTCGTGTGGGCGACATA	235
Db	62	GCACACAGCTCCCGTCGTGCAGCTTTTCCACGGCAGTAGCTTCGTGTGGGCGACATA	121
QY	236	TCGACACCCAGACGATATCCCGCGACGACCTCACTCTGGTGCCCTTCCAAAGCGGACG	295
Db	122	TCGACACCCAGACGATATCCCGCGAGGACCTCACTGGTGCCCTTCCAAAGCTGACG	181
QY	296	AGTACCGCAAGCTCGGTTCTCTCGGCTTCGCGGGGCTCCCATTCGCGGCGCTACCGGACG	355
Db	182	AGTACCGCAAGCTCGGCTCTCTTGTGCTTCGCGGGGCTCCCATTCGCGGCGCTACCGGACG	241
QY	356	CGTTCTGCTGCTCGGGTAGGAGAGTCTCTCCGCTACGCGCATATGTCGCGGAGCGCACT	415
Db	242	CGTTCTGCTGCTCGGGTAGGAGAGTCTACCGGCTACGCGCTGCTGCTGCGGAGCGCACT	301

QY	416	TGSGGTGCGGTTCTCTGCGCAGCACGCGCCCTCGCGCTTGSGGCGCGTGGCGCACGG	475
Db	302	TGGGGTGGCGGCTCTCCGCCGAGCACGGGCCCGCGGGCTTGGGGGCGGTGGAGCGCGG	361
QY	476	CCATTGTTGCGSAGGGCTACGCGCGCATCTTTTTTTCGCAACTCCGTGGCCACTGAGAG	535
Db	362	CCGTCGTGCGGGAGGGGCTACGCGGCGCACTCTTCGCGCAACTCCGTGGCCACTGGAGAG	421
QY	536	TGTACCCCTCGAGAGCTACAGGACGTTGGGGCTGGAGAGAGTGCAGACAGGGGATGTGG	595
Db	422	TGTACCCCTCGAGAGCTACAGGACACTGGAGCGTGGAGGAGAGTGCAGAGAGGGATGTGG	481
QY	596	TCACCGTGGACCTGTGTAACCTCCGTTTTTATTAACACACACTCTGGGCAAGAGTAGTAACG	655
Db	482	TCACCGTGGACCTGTGTAACCTCCGTTTTTATTAACACACACTCTGGGCAAGAGTAGTAACG	541
QY	656	TGAACCAATTGGTGTGATGTGGGCC	680
Db	542	TGAACCAATTGGTGTGATGTGATGTGC	565

RESULT 11	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM
BG047825	BG047825	OVL_28_c06.b1_A002 Ovary 1 (OVL) Sorghum bicolor cDNA, mRNA sequence.	BG047825	BG047825	GI:12497958	EST.	Sorghum. Sorghum bicolor

Eukaryota: Viridiplantae; Streptophyta: Embryophyta: Tracheophyta; Spermatophyta: Magnoliophyta: Liliopsida: Poales; Poaceae; PACC clade; Panicoideae; Andropogoneae; Sorghum.

REFERENCE 1 (bases 1 to 551)
Cordonier-Pratt, M.-M., Gingle, A., Marsala, C., Sudman, M. and Pratt, L.H.

TITLE An EST database from Sorghum: ovaries of varying immature stages
JOURNAL Unpublished (2000)
CONTACT: Cordonier-Pratt MM
LABORATORY for Genomics and Bioinformatics
The University of Georgia, Department of Plant Biology
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
Tel: 706 542 1860
Fax: 706 583 0210
Email: mmp@prattuga.edu

Sequences have been trimmed to exclude PolyA, vector and regions below Phred quality 16. The threshold for highest quality sequence is 20.

Seq primer: JEN REV
High quality sequence stop: 521
POLYA-No.

FEATURES
source Location/Qualifiers

1..551
/organism="Sorghum bicolor"
/db_xref="taxon:4558"
/clone_lib="Ovary 1 (OVI)"
/note="Organ: Mix of ovaries of varying immature stages from 8-week-old plants; Vector: pBluescript II from Lambda Zap II; Site_1: XhoI; Site_2: EcoRI; The library was made from poly-A RNA in the cloning vector Lambda Zap II. Clones to be sequenced were prepared by mass excision."
BASE COUNT 96 a 170 c 175 g 110 t
ORIGIN

Query Match 43.9%; Score 453; DB 12; Length 551;
Best Local Similarity 92.8%; Pred. No. 2.9e-63;
Matches 474; Conservative 1; Mismatches 36; Indels 0; Gaps 0;

254 TCCCGCGGAGACCTCACTGTGTCCTCAAGCGGAGAGTACCGAAGCTCGTT 313
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
9 TCCCGCGGAGACCTCACTGTGTCCTCAAGCGGAGAGTACCGAAGCTCGTT 68
314 CCTTGCCCTTGGCGGGCTCCCATCGCGGCTTACCGGAGCCGTTGTCGCGG 373
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
69 CCTTGCCCTTGGCGGGCTCCCATCGCGGCTTACCGGAGCCGTTGTCGCGG 128
374 AGGAGTCCCTCCGCTACGCTATGTCGCGGAGGCAACTTGGGCGGTTCTCTC 433
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
129 AGGAGTCCGCGCTACGCTATGTCGCGGAGGCAACTTGGGCGGTTCTCTC 188
434 GCGAGCAGCGCGCTGCGCTTGGGCGGCTGCGGAGCGGCACTTGTTCGAGGG 493
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
189 GCGAGCAGCGCGCTGCGCTTGGGCGGCTGCGGAGCGGCACTTGTTCGAGGG 248
494 AGCGCGGATCTTTTGGCAACTCCGTCGCGGCACTGGAGAGTGTACCTCTGA 553
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
249 AGCGCGGATCTTTTGGCAACTCCGTCGCGGCACTGGAGAGTGTACCTCTGA 308
554 CGGAGTGGGCGCTGGAAGAGTCAAGACAGGAGGATGTGTACCGTGGAGCTT 613
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
309 CGGAGACAGGAGCGCTGGAAGAGTCAAGACAGGAGGATGTGTACCGTGGAG 368
614 ACTCGGTTTTTATTACCACTCTGCGAAGAGTACAGCTGAACCAATTGGTAG 673
369 ACTCGGTTTTTATTACCACTCTGCGAAGAGTACAGCTGAACCAATTGGTAG 428
674 CTGGCCCTGTATTAGCGGAGGAGATCTTGGCTAGCGCCGGAAGACAGAGTAT 723
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
429 CTGGCCCTGTATTAGCGGAGGAGATCTTGGCTAGCGCTTGAAGACAGAGTAT 488
734 CGTCGAAAAGCTGCTCATGAGGAAAGCTTA 764
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
489 CGTCGAAAAGCTGCTCATGAGGAAAGCTTA 519

RESUME 12

LOCUS BE360264 546 bp mRNA linear EST 20-JUL-2000

DEFINITION DGI_62_B02.D1_A002 Dark grown 1 (DGI) Sorghum bicolor cDNA, mRNA

ACCESSION BE360264
VERSION BE360264.1 GI:9301821

KEYWORDS EST.

SOURCE sorghum.

ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoideae; Andropogoneae; Sorghum.

REFERENCE 1 (bases 1 to 546)
Cordonier-Pratt, M.-M., Gingle, A., Marsala, C., Sudman, M. and Pratt, L.H.

TITLE An EST database from Sorghum: dark-grown seedlings
JOURNAL Unpublished (2000)
CONTACT: Cordonier-Pratt MM
LABORATORY for Genomics and Bioinformatics
The University of Georgia, Department of Plant Biology
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
Tel: 706 542 1860
Fax: 706 583 0210
Email: mmp@prattuga.edu

Sequences have been trimmed to exclude PolyA, vector and regions below Phred quality 16. The threshold for highest quality sequence is 20.

Seq primer: JEN REV
High quality sequence stop: 512
POLYA-No.

FEATURES
source Location/Qualifiers

1..546
/organism="Sorghum bicolor"
/db_xref="taxon:4558"
/clone_lib="Dark grown 1 (DGI)"
/note="Organ: 5-day-old dark-grown seedlings; Vector: Lambda Zap; Site_1: XhoI; Site_2: EcoRI; The library was made from poly-A RNA in the cloning vector Lambda Zap II. Clones to be sequenced were prepared by mass excision."
BASE COUNT 70 a 220 c 162 g 94 t
ORIGIN

Query Match 42.8%; Score 442.6; DB 10; Length 546;
Best Local Similarity 91.6%; Pred. No. 1.4e-61;
Matches 479; Conservative 1; Mismatches 40; Indels 3; Gaps 1;

1 ATCATGGCGGCGCTCTGTGCGGAGCGGCTGTCCACGCGCTTCTAGCCCAATC 60
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
24 AACATGGCGGCTCTGTGCGGAGCGGCTGTCCACGCGGCTTCTAGCCCAATC 83
61 CGAGCTCCACAGCGGCTTATCCGCGGCTCCAGCTCACTGTCTACCTCCACTCA 120
DB 84 CGAGCTCCACAGCGGCTTATCCGCGGCTCCAGCTCACTGTCTACCTCCACTCA 143
121 CTAAATGCCCGCGGCTGTGCTCATGTCGCCGCGGCTGCTGCGCGGCGG--GC 177
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
144 CTAAATGCCCGCGGCTGTGCTCATGTCGCCGCGGCTGCTGCGCGGCGGCGG 203
DB 178 AGCACTCGCGCTGTGCTCATGTCGCCGCGGCTGCTGCGCGGCGGCGGCAATATC 237
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
204 AGCACTCGCGCTGTGCTCATGTCGCCGCGGCTGCTGCGCGGCGGCGGCAATATC 263
DB 238 GACACGACAGATATCCCGCGGAGACACTCACTGTGTCCTTCAAGCGGAGAG 297
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
264 GACACGACAGATATCCCGCGGAGACACTCACTGTGTCCTTCAAGCGGAGAG 323
DB 298 TACGCGAGCTGCTTCTGCTTGTGCGGCGGCTCCATCCGCGGCTTCCGAGCGCG 357
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
324 TACGCGAGCTGCTTCTGCTTGTGCGGCGGCTCCATCCGCGGCTTCCGAGCGCG 383

OY 358 TTCTGCTCCGCGGTGAGAGTCTCCCGCTACGCCATCTTCTCGCGGAGCACTTC 417
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 384 TTCTGCGCCCGGGGTGAGAGTCCACCGCTACGCCGCTGTCTGTGGGCAACTTC 443
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 418 GGTGCGGCTCTCTCTCCGAGCAAGCGCGCTGCGCTTGGGCGCCCTGCGGAGCGGC 477
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 444 GGTGCGGCTCTCTCTCCGAGCAAGCGCGCTGCGCTTGGGCGCCCTGCGGAGCGGC 503
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 478 ATTGTGCGAGGCTACGCGCGCATCTTTTTCGCACTCCG 520
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 504 GTCTGCGGAGGCTACGCGCGCATCTTCTTCGCACTCCG 546
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 13 841 bp mRNA linear EST 22-OCT-2001
BG309960
LOCUS HVSMEC0015H22f Hordeum vulgare seedling shoot EST library
DEFINITION HVSMEC0015H22f (Etisolated and unstressed) Hordeum vulgare cDNA clone
HVSMEC0015H22f, mRNA sequence.

ACCESSION BG309960
VERSION BG309960.1 GI:13110807
KEYWORDS EST.
SOURCE Hordeum vulgare.
ORGANISM Hordeum vulgare
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae
; Triticeae; Hordeum.
1 (bases 1 to 841)

REFERENCE Wing, R., Close, T.J., Kleinbols, A., Wise, R., Begum, D., Frisch, D., Yu
Y., Henry, D., Palmer, M., Rambo, T., Simmons, J., Choi, D.W., Fenton
R.D., Oates, R. and Main, D.
Development of a genetically and physically anchored EST resource
for barley genomics: Morex unstressed seedling shoot cDNA library
Unpublished (2001)

JOURNAL COMMENT Contact: Wing RA
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 7288
Fax: 864 656 4293
Email: rwing@clemson.edu
Total hg bases = 553
Seq primer: AATTACCTCTCACTAAGG
High quality sequence stop: 712.
Location/Qualifiers

FEATURES
source
1..841
/organism="Hordeum vulgare"
/cultivar="Morex"
/db_xref="taxon:4513"
/clone="HVSMEC0015H22f"
/clone_1b="Hordeum vulgare seedling shoot EST library
HVSMEC0003 (Etisolated and unstressed)"
/tissue-type="Seedling shoot"
/lab_host="TUC121"
/note="Vector: lambdaZAP: Site 1: EcoRI; Site 2: XhoI;
Seeds were surface sterilized then germinated under axenic
conditions in the dark at room temperature on filter paper
with water, nystatin and ceftaxime in covered
crystallization dishes. Five-day old seedling shoots were
then harvested, total RNA was prepared, poly(A) RNA was
purified, one primary unamplified cDNA library was made,
and 1 million pfu were in vivo excised to give plasmid
SK(-) cDNA phagemids. These steps were performed in the TJ
Close laboratory at the University of California,
Riverside (Choi, Close, Fenton). Phagemids were plated and
picked at the Clemson University Genomics Institute (CUGI)
(Begum, Palmer, Frisch, Atkins and Wing). Plasmid DNA
preparations, DNA sequencing and sequence analysis were
performed at CUGI (Wing, Yu, Frisch, Henry, Simmons, Oates
Rambo, Main). The sequence has been trimmed to remove
vector sequence and contains a minimum of 100 bases of
phred value 20 or above. For more details on library
preparation and sequence analysis see

http://www.genome.clemson.edu/projects/barley. To order
this clone see http://www.genome.clemson.edu/orders Also
see Close RJ, Wing R, Kleinbols A, Wise R (2001)
Genetically and physically anchored EST resources for
barley genomics. Barley Genetics Newsletter 31:29-30
(http://wheat.pw.usda.gov/g9pages/bgn/31/cover.html)"
BASE COUNT 136 a 308 c 247 g 146 t 4 others
ORIGIN

Query Match 42.1%; Score 435.4; DB 12; Length 841;
Best Local Similarity 77.0%; Pred. No. 1.5e-60;
Matches 558; Conservative 1; Mismatches 151; Indels 15; Gaps 2;

OY 22 GGAAGCGGCTGTCACGCGAGCGCTTCTAGCCCAATCCAGCTCAACAGCGCGTT 81
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 93 GCGCGAGCGAGAGCGAGCGCGCTGTCTAGCCCGCGCGCGCGCGCGCGCGCG 152
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 82 ATCCGCGGCTCCAGCTGACCTGTCATCGGCTCTCACTAAATGCGCGCGCTGG 141
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 153 ATTACG 209
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 142 TCCATGTCGCCCGCGCGCGCTGCTGCGCGCGCGCGCGCGCGCGCGCGCG 201
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 210 CCCCTAACCG 257
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 202 TTCCAGCGGAGCTCTCTGCGGCGGCGGACATATGACACGAGCATATCCCGCG 261
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 258 TTCCAGCGGAGCTCTCTGCGGCGGCGGACATATGACACGAGCATATCCCGCG 317
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 262 GAGCACTGATCTGCTGCTCCCAAGCGAGGAGTACCGAGCTCGCTCTCTCGCG 321
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 318 GAGCACTGATCTGCTGCTCCCAAGCGAGGAGTACCGAGCTCGCTCTCTCGCG 377
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 322 TTCCGCGGCTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 381
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 378 TTCCGCGGCTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 437
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 382 TCCGCTACGCGCATCTTCTGCGGCGGAGCGGAGTACCGGCTCTCTCTCGCG 441
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 438 TCAGCTACGCGCATCTTCTGCGGCGGAGCGGAGTACCGGCTCTCTCTCGCG 497
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 442 GCGCGCGCTGCTGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 501
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 498 GCGCGCGCTGCTGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 557
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 502 ATCTTTTTCGCACTCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 561
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 558 ATCTTTTTCGCACTCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 617
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 562 GGGCGCTGAGGAGTGTGACGAGGAGTGTGACGAGTGTGACGAGTGTGACG 621
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 618 GGTGCGAGGAGGAGTGTGACGAGGAGTGTGACGAGTGTGACGAGTGTGACG 677
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 622 TTTTAAACCACTCTGCGAGGAGTGTGACGAGTGTGACGAGTGTGACGAGT 681
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 678 TTGATTAACCACTCTGCGAGGAGTGTGACGAGTGTGACGAGTGTGACGAGT 737
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 682 GTATTGAGGCGGAGGAGTGTGCTTCTGAGCGCGGAGGAGGAGTGTGAGT 741
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 738 GNTATGAGGCGGAGGAGTGTGCTTCTGAGCGCGGAGGAGGAGTGTGAGT 797
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 742 GCTGC 746
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 798 TCTGC 802
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 14 534 bp mRNA linear EST 19-JUL-2000
AM747153
LOCUS WSI_66.G07_b1_A002 Water-stressed 1 (WS1) Sorghum bicolor cDNA,
DEFINITION mRNA sequence.
ACCESSION AM747153
VERSION AM747153.1 GI:7660891


```

Db      200 GGtGtCTCAAAATAGTTGCGGCTACCGAAATTAATGATGAATCATCAATTTGCTTT 141
Qy      902 GTCACAGATCGTTTTT---TTGTTACTAGTACTGTACAAATTGTACTCTGCTGCT 957
Db      140 GTCACAGATCGTTTTTGTGTTGTACTACTGTACAAATTGTACTCTGCTGCTGCT 81
Qy      958 ACTGTTCTATCTGTTGAATAACTGCTCTGTGCCA 994
Db      80 ACTGCTTATCTGTTGAATAACTGCTGTGTGCCA 44

```

Search completed: April 3, 2003, 04:15:21
 Job time : 1508 secs

GenCore version 5.1.4.p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 3, 2003, 02:59:38 ; Search time 75 Seconds
(without alignments)
4223.961 Million cell updates/sec

Title: US-10-027-450-46
Perfect score: 1033
Sequence: 1 atcatgcgcgcgcgtctc.....aaaaaaaaaaaaaaaa 1033

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents NA:*
1: /cgn2_6/ptodata/1/ina/5A.COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B.COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A.COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B.COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCBUS.COMB.seq:*
6: /cgn2_6/ptodata/1/ina/Backfile1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1032.6	100.0	1033	4	US-09-173-300-46
2	961.2	93.0	995	4	US-09-173-300-52
3	462.2	44.7	1112	4	US-09-173-300-48
4	291.2	28.2	1107	4	US-09-173-300-50
5	57.2	5.5	4403765	4	US-09-103-840A-2
6	57	5.5	1466	4	US-08-984-919A-10
7	57	5.5	1466	4	US-08-984-919A-12
8	57	5.5	1472	4	US-08-781-420-10
9	57	5.5	1472	4	US-08-781-420-12
10	57	5.5	1472	4	US-08-874-102-10
11	57	5.5	1472	4	US-08-874-102-12
12	57	5.5	1472	4	US-09-006-595A-10
13	57	5.5	1472	4	US-09-006-595A-12
14	57	5.5	1875	4	US-08-984-919A-46
15	57	5.5	1875	4	US-08-984-919A-48
16	57	5.5	1881	4	US-08-874-102-46
17	57	5.5	1881	4	US-08-874-102-48
18	55.4	5.4	1772	2	US-08-960-022-13
19	53.4	5.2	985	4	US-09-056-556-182
20	53.4	5.2	985	4	US-09-072-596-177
21	52	5.0	991	3	US-08-924-747-25
22	52	5.0	991	4	US-09-247-373B-25
23	52	5.0	991	4	US-09-296-715-25
24	51.8	5.0	1359	4	US-09-387-574-11
25	51.8	5.0	1359	4	US-09-668-096-11
26	51.6	5.0	3320	4	US-09-394-200-1
27	51.2	5.0	1534	1	US-08-300-903A-6

28	51.2	5.0	3080	4	US-09-099-041A-25	Sequence 25, Appl
29	51.2	5.0	3080	4	US-09-245-281-25	Sequence 25, Appl
30	51.2	5.0	3080	4	US-09-207-359B-25	Sequence 25, Appl
31	50.2	4.9	3437	3	US-08-860-339-17	Sequence 17, Appl
32	50.2	4.8	6476	4	US-09-127-670-5	Sequence 5, Appl
33	49.4	4.8	44377	2	US-08-804-227C-7	Sequence 7, Appl
34	49.4	4.8	44377	2	US-08-804-198-1	Sequence 1, Appl
35	49.2	4.8	1474	4	US-08-821-994-64	Sequence 64, Appl
36	48.8	4.7	1313	4	US-09-149-476-112	Sequence 112, App
37	48.8	4.7	1872	4	US-09-801-052-1	Sequence 1, Appl
38	48.8	4.7	2836	3	US-08-747-221B-24	Sequence 24, Appl
39	48.8	4.7	2836	3	US-08-747-221B-26	Sequence 26, Appl
40	48.8	4.7	2836	4	US-09-005-051-24	Sequence 24, Appl
41	48.8	4.7	2836	4	US-09-005-051-26	Sequence 26, Appl
42	48.8	4.7	3024	6	5284931-1	Patent No. 5284931
43	48.6	4.7	1098	3	US-09-248-335-35	Sequence 35, Appl
44	48.6	4.7	8920	2	US-08-446-855A-1	Sequence 1, Appl
45	48.6	4.7	8920	4	US-09-150-741-1	Sequence 1, Appl

ALIGNMENTS

```
RESULT 1
US-09-173-300-46
; Sequence 46, Application US/09173300
; Patent No. 6451581
; GENERAL INFORMATION:
; APPLICANT: Falco, Saverio Carl
; APPLICANT: Hitz, William D.
; APPLICANT: Kinney, Anthony J.
; APPLICANT: Kahn, Rebecca E.
; APPLICANT: Rafalski, J. Antoni
; TITLE OF INVENTION: PLANT BRANCHED CHAIN AMINO ACID BIOSYNTHETIC ENZYMES
; FILE REFERENCE: BB-1126
; CURRENT APPLICATION NUMBER: US/09/173,300
; EARLIER FILING DATE: 1998-10-15
; EARLIER APPLICATION NUMBER: 60/063,423
; EARLIER FILING DATE: 1997 October 28
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: Microsoft Word Version 7.0A
; SEQ ID NO 46
; LENGTH: 1033
; TYPE: DNA
; ORGANISM: Zea mays
US-09-173-300-46

Query Match      100.0%  Score 1032.6; DB 4; Length 1033;
Best Local Similarity 100.0%; Pred. No. 4.7e-226;
Matches 1033; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ATCATGCGCGGCGCTGTGCGGGAGCGGCGGTCTCAGCGAGCGCTTCTAGCCCAATC 60
    |||||
Db 1 ATCATGCGCGGCGCTGTGCGGGAGCGGCGGTCTCAGCGAGCGCTTCTAGCCCAATC 60

OY 61 CGAGCTCCAAACGAGCGCTTATTCGCGCGCTCCAGCTACCTGTATCGCTCCACTCA 120
    |||||
Db 61 CGAGCTCCAAACGAGCGCTTATTCGCGCGCTCCAGCTACCTGTATCGCTCCACTCA 120

OY 121 CTAAATGCGCGCGCGCTGTGCGGTCCATCGTCCCGGCGCGGTGCGCGGCGGAGCG 180
    |||||
Db 121 CTAAATGCGCGCGCGCTGTGCGGTCCATCGTCCCGGCGCGGTGCGCGGCGGAGCG 180

OY 181 AGCTGCGCGCTGTGCGGTTCATCGAGGCGAGTCTTGCTGTGTGGCGGAGCAATATCGAC 240
    |||||
Db 181 AGCTGCGCGCTGTGCGGTTCATCGAGGCGAGTCTTGCTGTGTGGCGGAGCAATATCGAC 240

OY 241 ACCGACAGATCATTCGCCCGGAGCACTCATCTGTGCGCTCCAGCGGAGAGATAC 300
    |||||
Db 241 ACCGACAGATCATTCGCCCGGAGCACTCATCTGTGCGCTCCAGCGGAGAGATAC 300

OY 301 CGCAGCTGCTTCTTCTGCGCTTTCGCGGCGGCTCCATCGCGGCGCTTACCGAGCGCTTC 360
    |||||
```

Db 301 CGCAAGCTGGTTCTTCGCTTCGGGGGCTCCCATCCGCGGCTCAACCGAGCGCGTTC 360
OY 361 GTGCTCCGGGTGAGAGTCTCCGCTACGCCATATTGTGGCGGAGCCAACTTCGGG 420
Db 361 GTGCTCCGGGTGAGAGTCTCCGCTACGCCATATTGTGGCGGAGCCAACTTCGGG 420
OY 421 TGGGGTTCCTTCGGCGAGCAGCGCGGCTTCGGGGCGCTGGGCGCAGCGGCATT 480
Db 421 TGGGGTTCCTTCGGCGAGCAGCGCGGCTTCGGGGCGCTGGGCGCAGCGGCATT 480
OY 481 GTTGCAGAGGCTACGCGCGCATCTTTTTCGCAACTCCGTGGCACTGAGAGGTATAC 540
Db 481 GTTGCAGAGGCTACGCGCGCATCTTTTTCGCAACTCCGTGGCACTGAGAGGTATAC 540
OY 541 CTTCTGAGCTCAGGACCTTTGGGGCTTGAAGAGTCAAGACAGAGGATGGTCTAC 600
Db 541 CTTCTGAGCTCAGGACCTTTGGGGCTTGAAGAGTCAAGACAGAGGATGGTCTAC 600
OY 601 GTGAGCTTGTCTAACTCCGTTTTTATTAACACACCTCTGGCAAGAGTACAAGCTGAAA 660
Db 601 GTGAGCTTGTCTAACTCCGTTTTTATTAACACACCTCTGGCAAGAGTACAAGCTGAAA 660
OY 661 CCAATTGTGATGCTGCCCTGTATATGAGCGGAGGATCTTTGCTACGCCGGAAG 720
Db 661 CCAATTGTGATGCTGCCCTGTATATGAGCGGAGGATCTTTGCTACGCCGGAAG 720
OY 721 ACAGAGTGAAGTGTGCTGAAAGCTGCTGACATGAGGAGGAACTTATGAGCGAGCTCTG 780
Db 721 ACAGAGTGAAGTGTGCTGAAAGCTGCTGACATGAGGAGGAACTTATGAGCGAGCTCTG 780
OY 781 CGGAGATGAAGAACTGAGTGTGAGTGAAGTCACTGACCTACTGATGTGCA 840
Db 781 CGGAGATGAAGAACTGAGTGTGAGTGAAGTCACTGACCTACTGATGTGCA 840
OY 841 CGGTGTCTCAAAATAAGTGGGCGCTACCGAAATATGAGTGAATCAATTTGCTT 900
Db 841 CGGTGTCTCAAAATAAGTGGGCGCTACCGAAATATGAGTGAATCAATTTGCTT 900
OY 901 TGTACAGATGCTTTTGTGTACTAGTACTGTACAAATTGACTCTGCGTACTACT 960
Db 901 TGTACAGATGCTTTTGTGTACTAGTACTGTACAAATTGACTCTGCGTACTACT 960
OY 961 GTTCTATCTGTGAATACCTGCTGTGCGCAAAAAAAAAAAAAAAAAAAAAAAAA 1020
Db 961 GTTCTATCTGTGAATACCTGCTGTGCGCAAAAAAAAAAAAAAAAAAAAAAAAA 1020
OY 1021 AAAAAAAAAAAAAA 1033
Db 1021 AAAAAAAAAAAAAA 1033

RESULT 2
US-09-173-300-52
Sequence 52, Application US/09173300
Patent No. 645181
GENERAL INFORMATION:
APPLICANT: Falco, Saverio Carl
APPLICANT: Hitz, William D.
APPLICANT: Kinney, Anthony J.
APPLICANT: Cahoon, Rebecca E.
APPLICANT: Rafalski, J. Antoni
TITLE OF INVENTION: PLANT BRANCHED CHAIN AMINO ACID BIOSYNTHETIC ENZYMES
FILE REFERENCE: BB-1126
CURRENT APPLICATION NUMBER: US/09/173.300
CURRENT FILING DATE: 1998-10-15
EARLIER APPLICATION NUMBER: 60/063.423
EARLIER FILING DATE: 1997 October 28
NUMBER OF SEQ ID NOS: 54
SOFTWARE: Microsoft Word Version 7.0A
SEQ ID NO 52
LENGTH: 995
TYPE: DNA
ORGANISM: Triticum aestivum

US-09-173-300-52
Query Match 93.0%; Score 961.2; DB 4; Length 995;
Best Local Similarly 99.0%; Pred. No. 8e-210;
Matches 978; Conservative 1; Mismatches 4; Indels 5; Gaps 1;

OY 26 CGGCGGTGTCACGCGCAGCGCTTCTAGCCCCCAATCCGAGCTCCACAGCGGCTTTATCC 85
Db 8 CGGCGGTGTCACGCGCAGCGCTTCTAGCCCCCAATCCGAGCTCCACAGCGGCTTTATCC 67
OY 86 GCGCTCCCACTACACTGTATGCGCTCCCACTACTAAATATGCGCGGCGGCTGGTCA 145
Db 68 GCGCTCCCACTACACTGTATGCGCTCCCACTACTAAATATGCGCGGCGGCTGGTCA 127
OY 146 TCGTCCCGCGCGCTCTCGCGCGGCGGCGAGAGTCCGCTGCTAGCGCTTTTCC 205
Db 128 TCGTCCCGCGCGCTCTCGCGCGGCGGCGAGAGTCCGCTGCTAGCGCTTTTCC 187
OY 206 ACGCGAGTCTCTGCTGTGGGCGCAATATGACACCGAGACAGATCATCCCGCGAGC 265
Db 188 ACGCGAGTCTCTGCTGTGGGCGCAATATGACACCGAGACAGATCATCCCGCGAGC 247
OY 266 ACTCTACTGTGTGCTCCCAAGCGGAGAGTACCGCAAGCTCGTTCCTTCCTTCG 325
Db 248 ACTCTACTGTGTGCTCCCAAGCGGAGAGTACCGCAAGCTCGTTCCTTCCTTCG 307
OY 326 CGGCGCTCCCATCCGCGCTTACCGGAGCGGCTTCGCTCCGCGGTGAGAGTCTCC 385
Db 308 CGGCGCTCCCATCCGCGCTTACCGGAGCGGCTTCGCTCCGCGGTGAGAGTCTCC 367
OY 386 GCTAGGCCATCATTTGCTCGCGGAGGCACTTGGGTGGGTTCTCTCGGAGCAGCGC 445
Db 368 GCTAGGCCATCATTTGCTCGCGGAGGCACTTGGGTGGGTTCTCTCGGAGCAGCGC 427
OY 446 CCGTCCGCTTGGGCGCGCTGCGCAGCGGCGCATTTGTCGAGGCGCTACGCGCATCT 505
Db 428 CCGTCCGCTTGGGCGCGCTGCGCAGCGGCGCATTTGTCGAGGCGCTACGCGCATCT 487
OY 506 TTTTTCGCACTCCGTGCGCACTGAGAGGTATACCTCTGAGAGTCAAGACGTTGGG 565
Db 488 TTTTTCGCACTCCGTGCGCACTGAGAGGTATACCTCTGAGAGTCAAGACGTTGGG 547
OY 566 CCGGAGAGAGTGCAGAGAGGAGTGTGTACCGGTGACCTTCTTAATCCGTTTTTA 625
Db 548 CCGGAGAGAGTGCAGAGAGGAGTGTGTACCGGTGACCTTCTTAATCCGTTTTTA 607
OY 626 TTAAACCACTCTGCGCAAGAGTACAACTGAAACCAATTGGTATGCTGCGCTGTA 685
Db 608 TTAAACCACTCTGCGCAAGAGTACAACTGAAACCAATTGGTATGCTGCGCTGTA 667
OY 686 TTGAGCGGAGGAGATCTTTGCTTACGCGCGGAGACAGAGATGATGGTGAAGCTG 745
Db 668 TTGAGCGGAGGAGATCTTTGCTTACGCGCGGAGACAGAGATGATGGTGAAGCTG 727
OY 746 CTGCATGAGGAA-----AGCTTATGACGCGAGCTCTGCGAGATGAAGATAGCT 800
Db 728 CTGCATGAGGAAAGATCAAGCTTATGACCGAGCTCTGCGAGATGAAGATAGCT 787
OY 801 GGAAGTAGAGTGAAGTATGATGATCAATCAATTTGCTTTGACAGATCGTTTTTTT 920
Db 788 GGAAGTAGAGTGAAGTATGATGATCAATCAATTTGCTTTGACAGATCGTTTTTTT 907
OY 861 CGGCTTACCGAAATATGATGATCAATCAATTTGCTTTGACAGATCGTTTTTTT 907
Db 848 CGGCTTACCGAAATATGATGATCAATCAATTTGCTTTGACAGATCGTTTTTTT 907
OY 921 TGTACTAGTACTGTACAAATTGTACTCGGCTGCTACTGTTCTTATCTGTTGAATA 980
Db 908 TGTACTAGTACTGTACAAATTGTACTCGGCTGCTACTGTTCTTATCTGTTGAATA 967
OY 981 CTGCTCTGTTGCCAAAAAAAAAAAAAAAA 1008
Db 968 CTGCTCTGTTGCCAAAAAAAAAAAAAAAA 995

```

RESULT 3
US-09-173-300-48
; Sequence 48, Application US/09173300
; Patent No. 6451581
; GENERAL INFORMATION:
; APPLICANT: Falco, Saverio Carl
; APPLICANT: Hiltz, William D.
; APPLICANT: Kinney, Anthony J.
; APPLICANT: Cahoon, Rebecca E.
; APPLICANT: Rafalski, J. Antoni
; TITLE OF INVENTION: PLANT BRANCHED CHAIN AMINO ACID BIOSYNTHETIC ENZYMES
; FILE REFERENCE: BB-1126
; CURRENT APPLICATION NUMBER: US/09/173.300
; EARLIER FILING DATE: 1998-10-15
; EARLIER APPLICATION NUMBER: 60/063,423
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: Microsoft Word Version 7.0A
; SEQ ID NO 48
; LENGTH: 1112
; TYPE: DNA
; ORGANISM: Oryza sativa
US-09-173-300-48

Query Match
Best Local Similarity 44.7%; Score 462.2; DB 4; Length 1112;
Matches 570; Conservative 1; Mismatches 149; Indels 9; Gaps 1;

27 GGGGCTGTCACGCGGCGCTTCTAGCCCAATCCGACCGACGCGGCTTATCCG 86
Db 72 GGGGGGGCGGGGAGACAGAGTTCTGACACCGGTCCCGCCCTGAGAGCTTCCGCG 131
QY 87 GGGCTCCACGCTACCTGTATCGCTCCCTCACTAAATGCGCGCGCTGGTCCAT 146
Db 132 CGGCACTGGTGGCGGCTATCTGCGCGCGCGCCCTGAAATGCCACGACATCTGCT 191
QY 147 C-----GTCCCGCGCGCGCTGCTGCGCGCGCGGCGGCGAGCTGCGCTGTCAGC 197
Db 192 GACCGCGCTGTCGCGCGCGCTGCGCGCGCTGCGCGCGCGGCGGAGCTGACGTGCGCG 251
QY 198 CGTTTTCACAGGCGAGTCTTCTGCTGGGCGACAAATATGACACCGACGATATCC 257
Db 252 CGATTTCACAGGCGAGTCTTCTGCTGGGCGATATGACACCGACGATATATCC 311
QY 258 CGCGGACGACCTACCTCTGCTGCTCCCTCCCAAGCGGAGATACGCAAGCTCGGTCCT 317
Db 312 GCGCGAGACCTGACCTGCTGCTCCCTCCCAAGCGGAGATACGCAAGCTCGGTCCT 371
QY 318 CGCTTCGCGGCGCTCCCATCCGCGCGCTACCCGAGCGCTTCTGCTCGGCTGAAGA 377
Db 372 CGCTTCGCTGCGCTCCCTCCCAAGCGGCGCTACCCGAGCGCTTCTGCTCGGCTGAAGA 431
QY 378 GTCTCCGCTACGCGCATATGTCGGGCGGAGCAACTTGGGTCGCTCTCTCCGCGA 437
Db 432 GACACACCGCTACGCGCTATATGTCGGGCGGCGCAACTTCCGCTGCTCTCCGCGA 491
QY 438 GCAAGCGCGCTGCGCTTGGGCGCTGCGCGACGCGCATTTGTCGSGAGGCTACGC 497
Db 492 GCAAGCGCGCTGCGCTTGGGCGCGCGCGCGCGCGCGCTGCTGCGCGCGAGGCTACGC 551
QY 498 GCGCATCTTTTTCGCACTGCTGCGCACTGGAGGATGATCCCTGCGCGCTACGCGA 557
Db 552 GCGCATCTTCTTCCGCACTGCTGCGCGCGCGCGCGCGCGCTGCTGCGCGCGAGCTACGCGA 611
QY 558 CGTTGGGCGCTGAGAGAGTGCACAAGAGGAGTGTGTCACCGTGAACCTTGTAACTC 617
Db 612 CACTGAGGCTGTGAAGAGTGCACAAGAGGAGTGTGTCACCGTGAACCTTGTAACTG 671
QY 618 CGTTTATTAACACACCTGTGGAAGAGTACAAAGCTGAACCAATTTGGTATGCTGG 677
Db 672 CGTATGATCAACACACATCCGGGAGAGTACAAAGCTGAAGCTTATCGCGGATGCGCG 731

```

```

QY 678 CCCTTAATTGAGCGGAGGATCTTTCCTTACGCCCGGAGACAGCAATGATTCGCTC 737
Db 732 GCGGCTATTGAGCGAGCGGAGATCTTTCCTTATGCCGGAAGACGGAATGATGCATC 791
QY 738 GAAGCTGC 746
Db 792 CAAGTCTGC 800

RESULT 4
US-09-173-300-50
; Sequence 50, Application US/09173300
; Patent No. 6451581
; GENERAL INFORMATION:
; APPLICANT: Falco, Saverio Carl
; APPLICANT: Hiltz, William D.
; APPLICANT: Kinney, Anthony J.
; APPLICANT: Cahoon, Rebecca E.
; APPLICANT: Rafalski, J. Antoni
; TITLE OF INVENTION: PLANT BRANCHED CHAIN AMINO ACID BIOSYNTHETIC ENZYMES
; FILE REFERENCE: BB-1126
; CURRENT APPLICATION NUMBER: US/09/173.300
; EARLIER FILING DATE: 1998-10-15
; EARLIER APPLICATION NUMBER: 60/063,423
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: Microsoft Word Version 7.0A
; SEQ ID NO 50
; LENGTH: 1107
; TYPE: DNA
; ORGANISM: Glycine max
US-09-173-300-50

Query Match
Best Local Similarity 28.2%; Score 291.2; DB 4; Length 1107;
Matches 430; Conservative 1; Mismatches 189; Indels 6; Gaps 2;

112 CTCACCTCCTAAATGCGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 171
Db 171 CCCAAGTCTCAACACCTTCCGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 230
QY 172 GCGGCGAGCAGCTGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 231
Db 231 TCCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 290
QY 232 AATATGACACGACGACATATCCCGCGAGACCTCACTGCTGCTGCTGCTGCTGCTGCT 291
Db 291 AATATGACACGACGACATATCCCGCGAGATCAAAACCAAGTACGCTGCTGCTGCTGCT 350
QY 292 GACGAGTACCGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 351
Db 351 GACGAGTACGAGAACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 407
QY 352 AGCGCGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 411
Db 408 AGCGCTTATGAAACCGCGGAGATCAAAACCAAGTACGCTGCTGCTGCTGCTGCTGCT 467
QY 412 AACTTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 471
Db 468 AACTTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 527
QY 472 CGCGCATGTTGGGAGGAGTACGCGCGCATTTTTCGCACTGCTGCTGCTGCTGCTGCT 531
Db 528 GCGCGAGTGTGCGGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 587
QY 532 GAGGTGTACCTCTTGGAGCTACGAGAGTGTGAGGAGTGTGAGGAGTGTGAGGAGTGT 591
Db 588 GAGGTGTATCCCTGAGAGTGTGAGGAGTGTGAGGAGTGTGAGGAGTGTGAGGAGTGT 644
QY 592 GTGTGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 651
Db 645 GTGTGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 704

```



```

COMPUTER: IBM PC compatible
OPERATING SYSTEM: Windows 95
SOFTWARE: WordPerfect for Windows, Version 7.0
CURRENT APPLICATION DATA: US/08/781,420
APPLICATION NUMBER: 1996
FILING DATE: December 3, 1996
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Verser, Carol Talkington
REGISTRATION NUMBER: 37,459
REFERENCE/DOCKET NUMBER: HW-2
TELECOMMUNICATION INFORMATION:
TELEPHONE: 970/493-7272
TELEFAX: 970/484-9505
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 1472 nucleotides
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 2..1105
US-08-781-420-10

Query Match          5.5%; Score 57; DB 4; Length 1472;
Best Local Similarity 66.9%; Pred. No. 0.00035;
Matches   81; Conservative    0; Mismatches 40; Indels    0; Gaps    0;

QY      913 TTTTGTGTTGGTAGACAGACTGTGTACAAATGTACTCCTGCCTGACTGTCTTAATCTGT 972
        ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db       1351 TTTTGTTTCCTTAGTGAAGTACTGTGTATTTGTATTAATACTATTTGCCCTGTTTTTAAAT 1410

QY      973 TTGAATACAGCTCGTGTCCCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1032
        ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db       1411 TTTTAATTAATTTTGTGCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1470

QY      1033 A 1033
Db       1471 A 1471

RESULT 9
US-08-781-420-12/C
Sequence 12, Application US/08781420
Patent No. 6248872
GENERAL INFORMATION:
APPLICANT: Chandrasekar, Ramaswamy
APPlicant: Mehla, Kapil
TITLE OF INVENTION: Parasitic Nematode Transglutaminase
TITLE OF INVENTION: Proteins, Nucleic Acid Molecules and
NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESS:
ADDRESSEE: Carol Talkington Verser, Ph.D.
ADDRESSEE: Heska Corporation
STREET: 1825 Sharp Point Drive
CITY: Fort Collins
STATE: Colorado
COUNTRY: USA
ZIP: 80525
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: Windows 95
SOFTWARE: WordPerfect for Windows, Version 7.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/781,420
FILING DATE: December 3, 1996
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Verser, Carol Talkington

```

```

? ; FEATURE: CDS
? ; NAME/KEY :
? ; LOCATION : 2..1105
US-08-874-102-10

Query Match          5.5% ; Score 57; DB 4; Length 1472;
Best Local Similarity 66.9% ; Pred No. 0.00035;
Matches      81; Conservative    0; Mismatches   40; Indels     0; Gaps     0;

OY      913 TTTTGTGTTAGTAGCTTTGTACAATGTACTCGTCCTGCTACTGTTCTTAICTGT 972
        ||||| ||| - | ||| - | ||| - | ||| - | ||| - | ||| - | ||| - |
DB      1351 TTTTGTTACTACTTANGTACTGTGTATTGTATTACTAATTTGGCCCTGTGTTTTAAAT 1410
OY      973 TTGAATAACCTCCTGCTGTGCCAAAAAaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 1032
        ||||| ||| - | ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      1411 TTTAAATAAATTTTTTTTGGAAAAAaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 1470
OY      1033 A 1033
        |
DB      1471 A 1471

RESULT 11
US-08-874-102-12/C
; Sequence 12, Application US/08874102
; Patent No. 6309644
GENERAL INFORMATION:
APPLICANT: Ramaswamy Chandrashekar
APPLICANT: Kapil Mehta
TITLE OF INVENTION: PARASITIC NEMATODE TRANSLUTAMINASE
TITLE OF INVENTION: PROTEINS, NUCLEIC ACID MOLECULES,
TITLE OF INVENTION: AND USES THEREOF
NUMBER OF SEQUENCES: 59
CORRESPONDENCE ADDRESS:
ADDRESSEE: Carol Talkington Verser, Ph.D.
ADDRESSEE: Heska Corporation
STREET: 1825 Sharp Point Drive
CITY: Fort Collins
STATE: Colorado
COUNTRY: USA
ZIP: 80525
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: Windows 95
SOFTWARE: Wordperfect for Windows, Version 7.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/874.102
FILING DATE: 12-JUN-1997
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Verser, Carol Talkington
REGISTRATION NUMBER: 37,459
REFERENCE/DOCKET NUMBER: HW-2-C1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 970/493-7272
TELEFAX: 970/484-9505
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 1472 nucleotides
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-08-874-102-12

Query Match          5.5% ; Score 57; DB 4; Length 1472;
Best Local Similarity 66.9% ; Pred. No. 0.00035;
Matches      81; Conservative    0; Mismatches   40; Indels     0; Gaps     0;

OY      913 TTTTGTGTTAGTAGCTTTGTACAATTTGTACTCCTGCCTGCTACTGTTCTTAICTGT 972
        ||||| ||| - | ||| - | ||| - | ||| - | ||| - | ||| - | ||| - |
DB      122 TTTGTGTACTTATGTTACTGTGTTAAATGTTATTAATTTTGGCCCTGTGTTTTAAAT 63

```



```
Patent No. 6383774
GENERAL INFORMATION:
APPLICANT: Chandrashekar, Ramaswamy
TITLE OF INVENTION: PARASITIC NEMATODE TRANSGLUTAMINASE
TITLE OF INVENTION: PROTEINS, NUCLEIC ACID MOLECULES,
NUMBER OF SEQUENCES: 59
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Carol Talkington Verser, Ph.D.
STREET: 1825 Sharp Point Drive
CITY: Fort Collins
STATE: Colorado
COUNTRY: USA
ZIP: 80525
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: Windows 95
SOFTWARE: Wordperfect for Windows, Version 7.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/984,919A
FILING DATE: 04-DEC-1997
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Verser, Carol Talkington
REGISTRATION NUMBER: 37,459
REFERENCE/DOCKET NUMBER: HW-2-C2
TELECOMMUNICATION INFORMATION:
TELEPHONE: 970/493-7272
TELEFAX: 970/484-9505
INFORMATION FOR SEQ ID NO: 46:
SEQUENCE CHARACTERISTICS:
LENGTH: 1875 nucleotides
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 24..1508
US-08-984-919A-46

Query Match          5.5%; Score 57; DB 4; Length 1875;
Best Local Similarity 66.9%; Pred. No. 0.00037;
Matches 81; Conservative 0; Mismatches 40; Indels 0; Gaps 0;

QY 913 TTTTGTGTTACTAGTACTGTACAATTGTAATTTACTCCTGCGCTACTGTTCTTATCTGT 972
    ||||| ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1754 TTTGTGTTACCTATGTAAGTGTGTATGTATGTAATTTTGCCTTTGTTTAAAT 1813
    || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 973 TTGAATACTGCTGTGTCGCCAATAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1032
    || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1814 TTTAATAAATTTTGTGGAATAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1873
    || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 1033 A 1033
DB 1874 A 1874

RESULT 15
US-08-984-919A-48/C
Sequence 48, Application US/08984919A
Patent No. 6383774
GENERAL INFORMATION:
APPLICANT: Chandrashekar, Ramaswamy
TITLE OF INVENTION: PARASITIC NEMATODE TRANSGLUTAMINASE
TITLE OF INVENTION: PROTEINS, NUCLEIC ACID MOLECULES,
NUMBER OF SEQUENCES: 59
CORRESPONDENCE ADDRESSES:
```

```
ADDRESSEE: Carol Talkington Verser, Ph.D.
ADDRESSEE: Heeka Corporation
STREET: 1825 Sharp Point Drive
CITY: Fort Collins
STATE: Colorado
COUNTRY: USA
ZIP: 80525
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: Windows 95
SOFTWARE: Wordperfect for Windows, Version 7.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/984,919A
FILING DATE: 04-DEC-1997
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Verser, Carol Talkington
REGISTRATION NUMBER: 37,459
REFERENCE/DOCKET NUMBER: HW-2-C2
TELECOMMUNICATION INFORMATION:
TELEPHONE: 970/493-7272
TELEFAX: 970/484-9505
INFORMATION FOR SEQ ID NO: 48:
SEQUENCE CHARACTERISTICS:
LENGTH: 1875 nucleotides
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-08-984-919A-48

Query Match          5.5%; Score 57; DB 4; Length 1875;
Best Local Similarity 66.9%; Pred. No. 0.00037;
Matches 81; Conservative 0; Mismatches 40; Indels 0; Gaps 0;

QY 913 TTTTGTGTTACTAGTACTGTACAATTGTAATTTACTCCTGCGCTACTGTTCTTATCTGT 972
    ||||| ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 122 TTTGTGTTACCTATGTAAGTGTGTATGTATGTAATTTTGCCTTTGTTTAAAT 63
    || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 973 TTGAATACTGCTGTGTCGCCAATAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1032
    || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 62 TTTAATAAATTTTGTGGAATAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 3
    || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 1033 A 1033
DB 2 A 2

Search completed: April 3, 2003, 04:39:51
Job time : 1467 secs
```

GenCore version 5.1.4.P5-4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 31, 2003, 09:04:48 ; Search time 36 Seconds
(without alignments)
921.650 Million cell updates/sec

Title: US-10-027-450-47
Sequence: 1266
1 MAALSGTAVSTAALAPIR.....AGGIFAYARRKTMASKAA 249

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues
Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: A.Geneseq-101002.*
2: /SID2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
3: /SID2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
4: /SID2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
5: /SID2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
6: /SID2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
7: /SID2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
8: /SID2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
9: /SID2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
10: /SID2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
11: /SID2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
12: /SID2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
13: /SID2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
14: /SID2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
15: /SID2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
16: /SID2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
17: /SID2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
18: /SID2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
19: /SID2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
20: /SID2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
21: /SID2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
22: /SID2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*
23: /SID2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1266	100.0	249	AA128439	Corn leud subunit
2	1226	96.8	244	AA128442	leud subunit of 3-
3	994.5	78.6	257	AA128440	Rice leud subunit
4	740.5	58.5	263	AA128441	leud subunit of 3-
5	710.5	56.1	251	AA128443	Arabidopsis thaila
6	710.5	56.1	251	AA128443	Arabidopsis thaila
7	710.5	56.1	251	AA128443	Arabidopsis thaila
8	710.5	56.1	251	AA128443	Arabidopsis thaila
9	710.5	56.1	251	AA128443	Arabidopsis thaila
10	651	51.4	253	AA128443	Herbicidally activ

11	644.5	50.9	256	23	AB192089	Herbicidally activ
12	627	49.5	256	23	AB192092	Herbicidally activ
13	627	49.5	256	23	AB192092	Herbicidally activ
14	258.5	20.4	164	22	AA196359	putative 3-isoprop
15	206.5	16.3	163	22	AA196359	Thermus thermophil
16	202	16.0	163	21	AA195395	Lactococcus lactis
17	191.5	15.1	191	15	AA194217	pseudomonas aerugi
18	191.5	15.1	191	23	AB194553	Listeria monocytog
19	184	14.5	212	22	AA196307	Staphylococcus epi
20	174.5	13.8	193	23	AB194817	Staphylococcus aur
21	166	13.1	189	23	AB195041	Amino acid sequenc
22	164.5	13.0	644	22	AA194234	Salmonella typhi c
23	162	12.8	164	18	AA192821	Staphylococcus aur
24	155.5	12.3	201	22	AA198423	E. coli cellular p
25	155	12.2	190	22	AA196564	Haemophilus influe
26	153.5	12.1	201	22	AA194427	C glutamicum prote
27	151	11.9	200	22	AA195572	C glutamicum prote
28	134	10.6	197	22	AA191199	C glutamicum prote
29	134	10.6	197	22	AA191199	C glutamicum prote
30	130	10.3	943	22	AA191445	C glutamicum prote
31	127.5	10.1	902	22	AB191080	Drosophila melanog
32	126	10.0	639	21	AA194440	Neisseria meningit
33	126	10.0	868	21	AA193301	Amino acid sequenc
34	125.5	9.9	553	22	AA195200	Corynebacterium gl
35	125.5	9.9	557	22	AA195200	Corynebacterium gl
36	125.5	9.9	639	21	AA194438	Neisseria meningit
37	125.5	9.9	869	21	AA193285	Amino acid sequenc
38	125	9.9	639	21	AA194438	Neisseria gonorrhoe
39	120	9.5	100	22	AA196055	Human isomerase-11
40	119.5	9.4	139	23	AB192235	Mitochondrial acou
41	115.5	9.1	780	20	AA198947	Amino acid sequenc
42	115.5	9.1	780	22	AA198433	S. epidermidis opé
43	114.5	9.0	160	22	AA191975	Lung cancer associ
44	112	8.8	417	21	AA198362	Mouse ischaemic co
45	112	8.8	889	23	AB197134	

ALIGNMENTS

RESULT 1	AA128439	AA128439 standard; Protein: 249 AA.
ID	AA128439	
AC	AA128439;	
XX		
DT	15-FEB-2000 (first entry)	
XX		
DE	Corn leud subunit of 3-isopropylmalate dehydratase amino acid sequence.	
XX		
KW	Corn; soybean; wheat; rice; dihydroxyacid dehydratase; leuc; leud;	
KW	branched chain amino acid transferase; biosynthetic enzyme; antibody;	
KW	3-isopropylmalate dehydratase.	
XX		
OS	Zea mays.	
XX		
PN	WO921880-A2.	
XX		
PD	06-MAY-1999.	
XX		
PF	20-OCT-1998; 98WO-US22081.	
XX		
PR	28-OCT-1997; 97US-0063423.	
XX		
PA	(DUPO) DU PONT DE NEMOURS & CO E. I.	
XX		
PI	Falco SC, Cahoon RE, Hiltz WD, Kinney AJ, Rafalski JA;	
XX		
DR	WPI; 2000-022904/02.	
XX		
XX	N-PSDB; AAX89462.	
PT	Nucleic acid fragments encoding branched chain amino acid biosynthetic enzymes	

XX Claim 16; Fig 7; 102pp; English.
 PS
 CC AAY28439-Y28442 are fragments of corn, soybean, wheat and rice leud
 CC subunit of 3-isopropylmalate dehydratase, amino acid sequences.
 CC Sequences AAY89442-X89465 are nucleic acid fragments that encode all or a
 CC substantial portion of dihydroxyacid dehydratase, a branched chain amino
 CC acid transferase, a leuc or a leud subunit of 3-isopropylmalate
 CC dehydratase from wheat, corn, soybean or rice. These enzymes are involved
 CC in biosynthesis and utilization of branched-chain amino acids. The
 CC nucleic acid sequences can be used to alter the level of expression of a
 CC branched chain amino acid biosynthetic enzyme in a host cell. They can
 CC also be used to obtain a nucleic acid fragment encoding a branched chain
 CC amino acid biosynthetic enzyme, and to identify inhibitors of a branched
 CC chain amino acid biosynthetic enzyme. The encoded proteins may be used
 CC to prepare antibodies for detecting the proteins in situ in cells, or in
 CC vitro in cell extracts.
 CC
 SQ Sequence 249 AA;
 Query Match 100.0%; Score 1266; DB 21; Length 249;
 Best Local Similarity 100.0%; Pred. No. 1.5e-119;
 Matches 249; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 1 MAAALSGTAVSTAAALAPRAPSFIIRRSQLTCHRLHSLKCRAGSIVPAAAAAGSS 60
 Db 1 MAAALSGTAVSTAAALAPRAPSFIIRRSQLTCHRLHSLKCRAGSIVPAAAAAGSS 60
 OY 61 SPSSAVFHGECEFYVDNIDTDQIIPAHLTLVPSKPDYRKLGSFAFAGLPSAAYPPFV 120
 Db 61 SPSSAVFHGECEFYVDNIDTDQIIPAHLTLVPSKPDYRKLGSFAFAGLPSAAYPPFV 120
 OY 121 AGESSSRATIVGANGCGSSREHAPVALGAGARAIYEGVARIFFPNSVATGCVTP 180
 Db 121 AGESSSRATIVGANGCGSSREHAPVALGAGARAIYEGVARIFFPNSVATGCVTP 180
 OY 181 LELTVGVAMKECKTGDDVTVDLANSVFINTHTSGKEYKLKPIGDAGPYIEAGGIFAVARKT 240
 Db 181 LELTVGVAMKECKTGDDVTVDLANSVFINTHTSGKEYKLKPIGDAGPYIEAGGIFAVARKT 240
 OY 241 GMTASKAA 249
 Db 241 GMTASKAA 249
 RESULT 2
 AAY28442
 ID AAY28442 standard; Protein: 244 AA.
 XX
 AC AAY28442;
 XX
 DT 15-FEB-2000 (first entry)
 XX
 DE Leud subunit of 3-isopropylmalate dehydratase amino acid sequence.
 XX
 KW Corn; soybean; wheat; rice; dihydroxyacid dehydratase; leuc; leud;
 KW branched chain amino acid transferase; biosynthetic enzyme; antibody;
 KW 3-isopropylmalate dehydratase.
 OS
 OS Triticum aestivum.
 XX
 PN WO9921880-A2.
 XX
 PD 06-MAY-1999.
 XX
 PF 20-OCT-1998; 98WO-US22081.
 XX
 PR 28-OCT-1997; 97US-0063423.
 XX
 PA (DUPO) DU PONT DE NEMOURS & CO E I.
 XX
 PI Falco SC, Cahoon RE, Hitz WD, Kinney AJ, Rafalski JA;

DR WPI: 2000-022904/02.
 DR N-PSDB: AAX89465.
 XX
 XX Nucleic acid fragments encoding branched chain amino acid biosynthetic
 PT enzymes
 PS
 CC Claim 16; Fig 7; 102pp; English.
 CC
 CC AAY28439-Y28442 are fragments of corn, soybean, wheat and rice leud
 CC subunit of 3-isopropylmalate dehydratase, amino acid sequences.
 CC Sequences AAY89442-X89465 are nucleic acid fragments that encode all or a
 CC substantial portion of dihydroxyacid dehydratase, a branched chain amino
 CC acid transferase, a leuc or a leud subunit of 3-isopropylmalate
 CC dehydratase from wheat, corn, soybean or rice. These enzymes are involved
 CC in biosynthesis and utilization of branched-chain amino acids. The
 CC nucleic acid sequences can be used to alter the level of expression of a
 CC branched chain amino acid biosynthetic enzyme in a host cell. They can
 CC also be used to obtain a nucleic acid fragment encoding a branched chain
 CC amino acid biosynthetic enzyme, and to identify inhibitors of a branched
 CC chain amino acid biosynthetic enzyme. The encoded proteins may be used
 CC to prepare antibodies for detecting the proteins in situ in cells, or in
 CC vitro in cell extracts.
 CC
 SQ Sequence 244 AA;
 Query Match 96.8%; Score 1226; DB 21; Length 244;
 Best Local Similarity 99.6%; Pred. No. 1.6e-115;
 Matches 240; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 OY 9 AVSTAAALAPRAPSFIIRRSQLTCHRLHSLKCRAGSIVPAAAAAGSSSPSSAVPH 68
 Db 4 AVSTAAALAPRAPSFIIRRSQLTCHRLHSLKCRAGSIVPAAAAAGSSSPSSAVPH 63
 OY 69 GECEFYVDNIDTDQIIPAHLTLVPSKPDYRKLGSFAFAGLPSAAYPPFVAPGESSR 128
 Db 64 GECEFYVDNIDTDQIIPAHLTLVPSKPDYRKLGSFAFAGLPSAAYPPFVAPGESSR 123
 OY 129 YAIIVGANGFCGSSREHAPVALGAGARAIYEGVARIFFPNSVATGCVTPLELDVGA 188
 Db 124 YAIIVGANGFCGSSREHAPVALGAGARAIYEGVARIFFPNSVATGCVTPLELDVGA 183
 OY 189 WKECKTGDDVTVDLANSVFINTHTSGKEYKLKPIGDAGPYIEAGGIFAVARKGMTASKAA 248
 Db 184 WKECKTGDDVTVDLANSVFINTHTSGKEYKLKPIGDAGPYIEAGGIFAVARKGMTASKAA 243
 OY 249 A 249
 Db 244 A 244
 RESULT 3
 AAY28440
 ID AAY28440 standard; Protein: 257 AA.
 XX
 AC AAY28440;
 XX
 DT 15-FEB-2000 (first entry)
 XX
 DE Rice leud subunit of 3-isopropylmalate dehydratase amino acid sequence.
 XX
 KW Corn; soybean; wheat; rice; dihydroxyacid dehydratase; leuc; leud;
 KW branched chain amino acid transferase; biosynthetic enzyme; antibody;
 KW 3-isopropylmalate dehydratase.
 OS
 OS Oryza sativa.
 XX
 PN WO9921880-A2.
 XX
 PD 06-MAY-1999.
 XX
 PF 20-OCT-1998; 98WO-US22081.
 XX
 PR 28-OCT-1997; 97US-0063423.

XX (DUPO) DU PONT DE NEMOURS & CO E I.
 XX Palco SC, Cahoon RE, Hitz WD, Kinney AJ, Rafalski JA;
 PI WPI: 2000-022904/02.
 DR N-PSDB: AAX89463.
 XX Nucleic acid fragments encoding branched chain amino acid biosynthetic
 PT enzymes -
 XX Claim 16, Fig 7, 102pp; English.
 XX AAY28439-Y28442 are fragments of corn, soybean, wheat and rice leud
 CC subunit of 3-isopropylmalate dehydratase, amino acid sequences.
 CC Sequences AAX89442-X89465 are nucleic acid fragments that encode all or a
 CC substantial portion of dihydroxyacid dehydratase, a branched chain amino
 CC acid transferase, a leuc or a leud subunit of 3-isopropylmalate
 CC in biosynthesis and utilization of branched-chain amino acids. The
 CC nucleic acid sequences can be used to alter the level of expression of a
 CC branched chain amino acid biosynthetic enzyme in a host cell. They can
 CC also be used to obtain a nucleic acid fragment encoding a branched chain
 CC amino acid biosynthetic enzyme, and to identify inhibitors of a branched
 CC chain amino acid biosynthetic enzyme. The encoded proteins may be used
 CC to prepare antibodies for detecting the proteins in situ in cells, or in
 CC vitro in cell extracts.
 XX Sequence 257 AA:
 SQ
 Query Match 78.6%; Score 994.5; DB 21; Length 257;
 Best Local Similarity 77.0%; Pred. No. 4,2e-92;
 Matches 198; Conservative 18; Mismatches 32; Indels 9; Gaps 2;
 QY 1 MAALSGTAVSTA-----ALAPRAPTSAPFIRRSQLTCHRLKLC--RRAGSTVPA 51
 DB 1 MAALAAAPALSTLEAPVAVLAPCPPTSPRTERRSRSWALICRPALKHNSPLTAAVA 60
 QY 52 AAAAAAGSSSPSSAVFGECEFYVGDNIPTDQIIPAEHLTVPSKPEDEYRKLSFAFAGLP 111
 DB 61 AAAAAAGBSTSAGVHGECEFYVGDNIPTDQIIPAEHLTVPSKPEDEYRKLSFAFAGLP 120
 QY 112 SAAYPPPEVAAGESSRYAIIIVGANGCGSSREHAPVALGAAGARAIVAEGRARIFERN 171
 DB 121 TAAYPPPEVAAGETTRVAVIIGANGCGSSREHAPVALGAAGARAIVAEGRARIFERN 180
 QY 172 SVATGEVYPLELTDVGAMKECKTGVDVTVDLANSVFINTSGEKYKLPIDGAGVIRAG 231
 DB 181 SVATGEVYPLELTDVGAMKECKTGVDVTVDELNCVAINHTSKQYKLPIDGAGVIRAG 240
 QY 232 GIRFAVARKTGMIASKAA 248
 DB 241 GIRFAVARKTGMIASKSA 257
 RESULT 4
 AAY28441
 ID AAY28441 standard; Protein: 263 AA.
 XX AAY28441;
 AC
 XX 15-FEB-2000 (first entry)
 DT
 XX Leud subunit of 3-isopropylmalate dehydratase amino acid sequence.
 DE
 XX
 KW Corn, soybean, wheat, rice, dihydroxyacid dehydratase; leuc, leud,
 KW branched chain amino acid transferase, biosynthetic enzyme; antibody;
 KW 3-isopropylmalate dehydratase.
 XX
 OS Glycine max.
 XX
 FT Key Location/Qualifiers
 FT Misc-difference 4

FT /label= unknown
 FT /note= "encoded by TAG"
 XX W09921880-A2.
 XX
 XX 06-MAY-1999.
 PD
 XX 20-OCT-1998; 98WO-US22081.
 PF
 XX 28-OCT-1997; 97US-0063423.
 PR
 XX (DUPO) DU PONT DE NEMOURS & CO E I.
 XX Palco SC, Cahoon RE, Hitz WD, Kinney AJ, Rafalski JA;
 PI WPI: 2000-022904/02.
 DR N-PSDB: AAX89464.
 XX Nucleic acid fragments encoding branched chain amino acid biosynthetic
 PT enzymes -
 XX Claim 16, Fig 7, 102pp; English.
 XX AAY28439-Y28442 are fragments of corn, soybean, wheat and rice leud
 CC subunit of 3-isopropylmalate dehydratase, amino acid sequences.
 CC Sequences AAX89442-X89465 are nucleic acid fragments that encode all or a
 CC substantial portion of dihydroxyacid dehydratase, a branched chain amino
 CC acid transferase, a leuc or a leud subunit of 3-isopropylmalate
 CC dehydratase from wheat, corn, soybean or rice. These enzymes are involved
 CC in biosynthesis and utilization of branched-chain amino acids. The
 CC nucleic acid sequences can be used to alter the level of expression of a
 CC branched chain amino acid biosynthetic enzyme in a host cell. They can
 CC also be used to obtain a nucleic acid fragment encoding a branched chain
 CC amino acid biosynthetic enzyme, and to identify inhibitors of a branched
 CC chain amino acid biosynthetic enzyme. The encoded proteins may be used
 CC to prepare antibodies for detecting the proteins in situ in cells, or in
 CC vitro in cell extracts.
 XX Sequence 263 AA:
 SQ
 Query Match 58.5%; Score 740.5; DB 21; Length 263;
 Best Local Similarity 62.8%; Pred. No. 2e-66;
 Matches 157; Conservative 24; Mismatches 52; Indels 17; Gaps 5;
 QY 8 TAVTALLAPRAPTSAPFIRRSQLTCHRL-----HSIKCRAGSIYPAAAAAA 56
 DB 20 TRSSAATVLP---RNLAFTKLSLSHSHLDPFLSFPPTPKSNPRNRVAVSLQTPRAGS 76
 QY 57 AGSSSPSSAVFGECEFYVGDNIPTDQIIPAEHLTVPSKPEDEYRKLSFAFAGPSAAY 116
 DB 77 AASASP-SASFHGLCYVGDNIPTDQIIPAEHLTVPSKPEDEYRKLSYALIGLP-ATYA 134
 QY 117 TPFVAPGEESRYAIIIVGANGCGSSREHAPVALGAAGARAIVAEGRARIFERNVATG 176
 DB 135 TRFIEPGEIKTKYAVIIGANGCGSSREHAPVALGAAGARAIVAEGRARIFERNVATG 194
 QY 177 EYVPLELTDVGAMKECKTGVDVTVDLANSVFINTSGEKYKLPIDGAGVIRAGIFAY 236
 DB 195 EYVPLE-SEGRICECTGVDVTVIELGSRILNHTGKRYKLPIDGAGVIRAGIFAY 253
 QY 237 ARKTGMISK 246
 DB 254 ARKTGMIPSR 263
 RESULT 5
 AAG26403
 ID AAG26403 standard; Protein: 251 AA.
 XX AAG26403;
 AC
 XX 17-OCT-2000 (first entry)
 DT
 XX

DE Arabidopsis thaliana protein fragment SEQ ID NO: 30845.
XX Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
XX termination sequence.
OS Arabidopsis thaliana.
XX EP1033405-A2.
XX
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136382.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139452.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139765.
PR 21-JUN-1999; 99US-0139817.

PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145182.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145224.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 05-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.

PR	30-AUG-1999.	99US-0151303.
PR	31-AUG-1999.	99US-0151438.
PR	01-SEP-1999.	99US-0151930.
PR	07-SEP-1999.	99US-0152363.
PR	10-SEP-1999.	99US-0153070.
PR	13-SEP-1999.	99US-0153758.
PR	15-SEP-1999.	99US-0154018.
PR	16-SEP-1999.	99US-0154039.
PR	20-SEP-1999.	99US-0154779.
PR	22-SEP-1999.	99US-0155139.
PR	23-SEP-1999.	99US-0155486.
PR	24-SEP-1999.	99US-0155659.
PR	28-SEP-1999.	99US-0156458.
PR	29-SEP-1999.	99US-0156596.
PR	04-OCT-1999.	99US-0157117.
PR	05-OCT-1999.	99US-0157753.
PR	06-OCT-1999.	99US-0157865.
PR	07-OCT-1999.	99US-0158029.
PR	08-OCT-1999.	99US-0158232.
PR	12-OCT-1999.	99US-0158369.
PR	13-OCT-1999.	99US-0159293.
PR	13-OCT-1999.	99US-0159294.
PR	13-OCT-1999.	99US-0159295.
PR	14-OCT-1999.	99US-0159329.
PR	14-OCT-1999.	99US-0159330.
PR	14-OCT-1999.	99US-0159331.
PR	14-OCT-1999.	99US-0159637.
PR	14-OCT-1999.	99US-0159638.
PR	18-OCT-1999.	99US-0159584.
PR	21-OCT-1999.	99US-0160741.
PR	21-OCT-1999.	99US-0160767.
PR	21-OCT-1999.	99US-0160768.
PR	21-OCT-1999.	99US-0160770.
PR	21-OCT-1999.	99US-0160814.
PR	21-OCT-1999.	99US-0160815.
PR	22-OCT-1999.	99US-0160980.
PR	22-OCT-1999.	99US-0160981.
PR	22-OCT-1999.	99US-0160989.
PR	25-OCT-1999.	99US-0161404.
PR	25-OCT-1999.	99US-0161405.
PR	25-OCT-1999.	99US-0161406.
PR	26-OCT-1999.	99US-0161359.
PR	26-OCT-1999.	99US-0161360.
PR	26-OCT-1999.	99US-0161361.
PR	28-OCT-1999.	99US-0161920.
PR	28-OCT-1999.	99US-0161992.
PR	28-OCT-1999.	99US-0161993.
PR	29-OCT-1999.	99US-0162142.

[illegible]

	RESULT	6
ID	AAG42502	
XX	AAG42502 standard; Protein: 251 AA.	
AC	AAG42502;	
XX		
DT	18-OCT-2000 (first entry)	
XX		
DE	Arabidopsis thaliana protein fragment SEQ ID NO: 53014.	
KW	Protein identification; signal transduction pathway; metabolic pathway;	
KM	hybridisation assay; genetic mapping; gene expression control; promoter;	
XX	termination sequence.	
XX		
OS	Arabidopsis thaliana.	
PN	EPI033405-AZ.	
XX		
PD	06-SEP-2000.	
XX		
PE	25-FEB-2000; 2000EP-0301439.	
XX		
PR	25-FEB-1999; 99US-0121825.	
PR	05-MAR-1999; 99US-0123180.	
PR	09-MAR-1999; 99US-0123548.	
PR	23-MAR-1999; 99US-0125788.	
PR	25-MAR-1999; 99US-0126264.	
PR	29-MAR-1999; 99US-0126785.	
PR	01-APR-1999; 99US-0127462.	
PR	06-APR-1999; 99US-0128234.	
PR	08-APR-1999; 99US-0128714.	
PR	16-APR-1999; 99US-0129845.	
PR	19-APR-1999; 99US-0130077.	
PR	21-APR-1999; 99US-0130449.	
PR	23-APR-1999; 99US-0130510.	
PR	23-APR-1999; 99US-0130891.	
PR	28-APR-1999; 99US-0131448.	
PR	30-APR-1999; 99US-0132048.	
PR	30-APR-1999; 99US-0132407.	
PR	04-MAY-1999; 99US-0132484.	
PR	05-MAY-1999; 99US-0132485.	
PR	06-MAY-1999; 99US-0132486.	
PR	06-MAY-1999; 99US-0132487.	
PR	07-MAY-1999; 99US-0132863.	
PR	11-MAY-1999; 99US-0132863.	
PR	14-MAY-1999; 99US-0134218.	
PR	14-MAY-1999; 99US-0134219.	
PR	14-MAY-1999; 99US-0134221.	
PR	14-MAY-1999; 99US-0134370.	
PR	18-MAY-1999; 99US-0134768.	
PR	19-MAY-1999; 99US-0134941.	
PR	20-MAY-1999; 99US-0135124.	
PR	21-MAY-1999; 99US-0135353.	
PR	24-MAY-1999; 99US-0135629.	
PR	25-MAY-1999; 99US-0136021.	
PR	27-MAY-1999; 99US-0136392.	
PR	28-MAY-1999; 99US-0136783.	
PR	01-JUN-1999; 99US-0137222.	
PR	03-JUN-1999; 99US-0137528.	
PR	04-JUN-1999; 99US-0137502.	
PR	07-JUN-1999; 99US-0137724.	
PR	08-JUN-1999; 99US-0138094.	
PR	10-JUN-1999; 99US-0138540.	
PR	10-JUN-1999; 99US-0138847.	
PR	14-JUN-1999; 99US-0139119.	
PR	16-JUN-1999; 99US-0139452.	
PR	16-JUN-1999; 99US-0139453.	
PR	17-JUN-1999; 99US-0139452.	
PR	18-JUN-1999; 99US-0139454.	
PR	18-JUN-1999; 99US-0139455.	
PR	18-JUN-1999; 99US-0139456.	
PR	18-JUN-1999; 99US-0139457.	

PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 24-JUN-1999; 99US-0140354.
PR 28-JUN-1999; 99US-0140685.
PR 29-JUN-1999; 99US-0140823.
PR 30-JUN-1999; 99US-0140991.
PR 01-JUL-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 02-JUL-1999; 99US-0142154.
PR 06-JUL-1999; 99US-0142055.
PR 08-JUL-1999; 99US-0142390.
PR 09-JUL-1999; 99US-0142803.
PR 12-JUL-1999; 99US-0142920.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144684.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147182.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147346.
PR 09-AUG-1999; 99US-0147493.
PR 10-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148564.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.

PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151086.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151920.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 05-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 56.1%; Score 710.5; DB 21; Length 251;
Best Local Similarity 59.5%; Pred. No. 2,1e-63;
Matches 153; Conservative 27; Mismatches 62; Indels 15; Gaps 6;

OY 1 MAALSGTAVSTALALPIRAPTSATFIRNSQLCHRLHSJKCRAGSIYMAAAAAGSS 60
DB 1 MAASIQSANPTLSKTLSPKPSFATFRKSPFL--KENSISV--ASNFRPLYSREASSSF 56
OY 61 SPSSA-----VHGECFVVGNDIDTDQITPAEHLLTVPSKPDXYRKJGSFAAGLP 113
DB 57 VTRSAAEPOEKRTFPHGLCYVVDNIDTDQITPAEFLTVPSNDEYEKIGSYALVGLP-A 115
OY 114 AYPPEFVAPGESSRVAIVYIGGANGFCSSREHAPVALGAGARATVAGYARIPFRNSV 173
DB 116 SYKERFVOPGEMKTKYISIIITIGENFGCGSSREHAPVCLGAGAKKAVVAOSYAIRFRNSV 175

PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 13-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 22-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 23-JUL-1999; 99US-0145226.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.

PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 20-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 23-AUG-1999; 99US-0150566.
PR 25-AUG-1999; 99US-0150884.
PR 26-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 27-AUG-1999; 99US-0151303.
PR 30-AUG-1999; 99US-0151308.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 21-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 56.1%; Score 710.5; DB 21; Length 261;
Best Local Similarity 59.5%; Pred. No. 2,2e-63;
Matches 153; Conservative 27; Mismatches 62; Indels 15; Gaps 6;

QY 1 MAALSTAVSTALALAPIRAPISAFIRRSQLTCHRLHSIKCRAGSIYPAAAAAGSS 60
Db 11 MAASLOSANPTLSRTLASPNKPSFATFRSPFL--RFNSTSV--ASNRPPLYSREASSSF 66

PR	04-AUG-1999	9905-0427204
PR	04-AUG-1999	9905-0473032
PR	03-AUG-1999	9905-0447192
PR	05-AUG-1999	9905-0447260
PR	06-AUG-1999	9905-0477303
PR	06-AUG-1999	9905-0447416
PR	06-AUG-1999	9905-0447493
PR	09-AUG-1999	9905-0479335
PR	10-AUG-1999	9905-0448171
PR	11-AUG-1999	9905-0448319
PR	12-AUG-1999	9905-0448341
PR	13-AUG-1999	9905-0448365
PR	13-AUG-1999	9905-0448684
PR	16-AUG-1999	9905-0449368
PR	17-AUG-1999	9905-0449175
PR	18-AUG-1999	9905-0449426
PR	20-AUG-1999	9905-0449722
PR	20-AUG-1999	9905-0449723
PR	20-AUG-1999	9905-0449929
PR	23-AUG-1999	9905-0449902
PR	23-AUG-1999	9905-0449930
PR	25-AUG-1999	9905-0505866
PR	26-AUG-1999	9905-0510884
PR	27-AUG-1999	9905-0515065
PR	27-AUG-1999	9905-0515066
PR	27-AUG-1999	9905-0515080
PR	30-AUG-1999	9905-0515133
PR	31-AUG-1999	9905-0515148
PR	01-SEP-1999	9905-0515293
PR	07-SEP-1999	9905-0515363
PR	10-SEP-1999	9905-0513070
PR	13-SEP-1999	9905-0537378
PR	15-SEP-1999	9905-05154018
PR	16-SEP-1999	9905-05154039
PR	20-SEP-1999	9905-05154779
PR	22-SEP-1999	9905-05155139
PR	23-SEP-1999	9905-05155486
PR	24-SEP-1999	9905-05155486
PR	28-SEP-1999	9905-0515658
PR	29-SEP-1999	9905-05156586
PR	04-OCT-1999	9905-05171117
PR	05-OCT-1999	9905-05175753
PR	06-OCT-1999	9905-0517865
PR	07-OCT-1999	9905-0518029
PR	08-OCT-1999	9905-0518823
PR	12-OCT-1999	9905-0518369
PR	13-OCT-1999	9905-0519293
PR	13-OCT-1999	9905-0519294
PR	13-OCT-1999	9905-0519329
PR	14-OCT-1999	9905-0519330
PR	14-OCT-1999	9905-0519331
PR	14-OCT-1999	9905-0519637
PR	18-OCT-1999	9905-0519584
PR	21-OCT-1999	9905-05160741
PR	21-OCT-1999	9905-05160767
PR	21-OCT-1999	9905-0516768
PR	21-OCT-1999	9905-0516770
PR	21-OCT-1999	9905-0516814
PR	22-OCT-1999	9905-05160815
PR	22-OCT-1999	9905-05160980
PR	22-OCT-1999	9905-05160981
PR	22-OCT-1999	9905-05160989
PR	25-OCT-1999	9905-05161404
PR	25-OCT-1999	9905-05161405
PR	26-OCT-1999	9905-05161406
PR	26-OCT-1999	9905-05161359
PR	26-OCT-1999	9905-05161361
PR	26-OCT-1999	9905-05161920
PR	28-OCT-1999	9905-05161993
PR	28-OCT-1999	9905-05161993

PR	29-OCT-1999:	99US-0162142.
Query Match	56.1%;	Score 710.5; DB 21; Length 261;
Best Local Similarity	59.5%;	Pred. No. 2.2e-63;
Matches 153; Conservative	27; Mismatches 62;	Indels 15; Gaps 6;
OY	1	MAALISGTAASVTALLAPIRAPTSAFIRRSOLTCRHLSLKCRRAGSIVPAAAAAGSS 60
DB	11	MAAISOGANPTLSRLTSLASPKNKPSFATREFSPPL--RFNSTSV--ASNFKPLVSRNASSSF 66
OY	61	SPSSA-----VFHGECFYVDNIDITDQIIPAEHLTLTPSKPDPEYRKLSPFAAGLPSPA 113
DB	67	VTRSAAPFOERKTFHGCICVYGDMIDTDQILPAEFLTLVPENPEYEKLSYALVGLP-A 125
OY	114	AYPEPFAPAEESSRRAIIVGANGFCGSSREHAVALAGAARAIVAEGYARIEFRNSV 173
DB	126	SYKEREPQPMEMTKTSIIIGENFECCSSREHAHPVCGAAGAKAVAAQSARIFEFRNSV 185
OY	174	ATGVVYPLELTDYGAKCKETGTGYTVYDL--ANSYFIWHTSGXEXKLPFGDGAPVTEAG 231
DB	186	ATGEVYPLD--SEVRVDCDECTGDVAIVLEEDSTILINHTTGKEYKLPFGDAGPVLDAG 244
OY	232	GIFAYARKTGMIASKAA 248
DB	245	GIFAYARKKGMTPSAA 261
RESULT 10		
ID	ABB92729	
AB	ABB92729 standard; Protein; 253 AA.	
XX	ABB92729;	
DX	31-MAY-2002 (first entry)	
DE	Herbicidally active polypeptide SEQ ID NO 1940.	
XX	Herbicidal; plant; agriculture; herbicide.	
OS	Arabidopsis thaliana.	
PN	WO200210210-A2.	
XX	07-FEB-2002.	
PD	28-AUG-2001; 2001WO-EP09892.	
PF	28-AUG-2001; 2001WO-EP09892.	
XX	28-AUG-2001; 2001WO-EP09892.	
PR	(FARB) BAYER AG.	
PA	Tietjen K, Weidler M;	
PI	WPI; 2002-269010/31.	
DR		
XX	Identifying plant target proteins for herbicidally active compounds,	
PT	comprising aligning and comparing nucleic acid or amino acid sequences	
PT	from plant with nucleic acid or amino acid sequences from non-plant	
PT	organisms -	
XX		
PS	Claim 5; SEQ ID NO 1940; 261pp + Sequence Listing; English.	
XX		
CC	The invention relates to identifying target proteins	
CC	(ABB90790-ABB94016) for herbicidally active compounds, comprising	
CC	aligning and comparing nucleic acid or amino acid sequences from plant	
CC	with nucleic acid or amino acid sequences from non-plant organisms using	
CC	suitable search parameters, where plant sequences having an E-value	
CC	greater by a factor of 3 than the E-value of most similar non-plant	
CC	sequences are selected. The polypeptides or nucleic acids encoding them	
CC	are useful for identifying modulators. The identified modulators are	
CC	useful as herbicides.	
XX		
Sequence	253 AA;	

Query Match 51.4%; Score 651; DB 23; Length 253;
 Best Local Similarity 59.7%; Pred. NO. 2.2e-57;
 Matches 135; Conservative 29; Mismatches 56; Indels 6; Gaps 4;

QY 24 SAIR-RSOLTCRHLSTKCRAGSIVPAAAAAGSSSPSSAVFHGECFVGNIDTDQ 82
 Db 29 SPFLQKASATIFENKPLTSSATITRYAASSSDSGESTIREPHGLCFYLNIDTDQ 88
 QY 83 ITAEHLTLVPSKPDYKRLGSFAFAGLPSAAYFPFPAPEESRYALITYGANEGCGS 142
 Db 89 ITAEHTLPSIPEDRELGSFALNGLPK-FYNERFVVPPEMKSYSVITIGDNEGCGS 147
 QY 143 SREHAPVALGAAGARAIVAGYARIFERNVSATGEVPLELTDVGAMKECKTGVDVYDL 202
 Db 148 SRHAPVCLGAAGAKAVAEYARIFERNVCATGEIFPLE-SEVARIOCECKTGVDVYIEH 206
 QY 203 ---ANSVFINHSGKEYKLPKIGDAGPYEAGGIFAYARKTGMTAS 245
 Db 207 KEDGSSLLNHTTRKEYKLKPLGDAGPVIDAGGIFAYARKAGMTIPS 252

RESULT 11

ABB92089
 ID ABB92089 standard; Protein; 256 AA.

AC ABB92089;

DT 31-MAY-2002 (first entry)

DE Herbicidally active polypeptide SEQ ID NO 1300.

KW Herbicidal; plant; agriculture; herbicide.

OS Arabidopsis thaliana.

PN W0200210210-A2.

PD 07-FEB-2002.

PF 28-AUG-2001; 2001WO-EP09892.

PR 28-AUG-2001; 2001WO-EP09892.

PA (FARB) BAYER AG.

PI Tietjen K, Weidler M;

DR WPI; 2002-269010/31.

PT Identifying plant target proteins for herbicidally active compounds,
 comprising aligning and comparing nucleic acid or amino acid sequences
 from plant with nucleic acid or amino acid sequences from non-plant
 organisms -

PS Claim 5; SEQ ID NO 1300; 261pp + Sequence Listing; English.

CC The invention relates to identifying target proteins
 CC (ABB90790-ABB94016) for herbicidally active compounds, comprising
 CC aligning and comparing nucleic acid or amino acid sequences from plant
 CC with nucleic acid or amino acid sequences from non-plant organisms using
 CC suitable search parameters, where plant sequences having an E-value
 CC greater by a factor of 3 than the E-value of most similar non-plant
 CC sequences are selected. The polypeptides or nucleic acids encoding them
 CC are useful for identifying modulators. The identified modulators are
 CC useful as herbicides.

SQ Sequence 256 AA;

Query Match 50.9%; Score 644.5; DB 23; Length 256;
 Best Local Similarity 57.9%; Pred. NO. 1e-56;
 Matches 139; Conservative 27; Mismatches 59; Indels 15; Gaps 6;

QY 1 MAAISGTAIVSTAALLAPIRAPTSAFIRRSOLTCRHLSTKCRAGSIVPAAAAAGSS 60
 Db 1 MAASIQSANPTLSRLTASPNKPSRATFRPFL--RNRSTV--ASNFKPLVSEASSSF 56
 QY 61 SPSSA-----VHGECEPVGDNIDTDQIIPAEHLTVPSKPDYKRLGSFAFAGLPSA 113
 Db 57 VTRSAAEPOERKTFHGLCYVVGDNIDTDQIIPAEFLTVPSPEYERKLSYALVGLP-A 115
 QY 114 AYTPFPVAPGEESRYALITYGANFGCGSSREHAPVALGAAGARAIVAGYARIFERNV 173
 Db 116 STKREVPDGEKRTKTSITIGENFGCGSSREHAPVCLGAAGAKAVAGSTARIFERNV 175
 QY 174 AGEVYVPLELTDVGAMKECKTGVDVYDL--ANSVFINHSGKEYKLPKIGDAGPYEAG 231
 Db 176 AGEVYVPLD-SEVARIOCECKTGVDVYIEH--SEVARIOCECKTGVDVYIEH 234

RESULT 12

ABB92090
 ID ABB92090 standard; Protein; 256 AA.

AC ABB92090;

DT 31-MAY-2002 (first entry)

DE Herbicidally active polypeptide SEQ ID NO 1301.

KW Herbicidal; plant; agriculture; herbicide.

OS Arabidopsis thaliana.

PN W0200210210-A2.

PD 07-FEB-2002.

PF 28-AUG-2001; 2001WO-EP09892.

PR 28-AUG-2001; 2001WO-EP09892.

PA (FARB) BAYER AG.

PI Tietjen K, Weidler M;

DR WPI; 2002-269010/31.

PT Identifying plant target proteins for herbicidally active compounds,
 comprising aligning and comparing nucleic acid or amino acid sequences
 from plant with nucleic acid or amino acid sequences from non-plant
 organisms -

PS Claim 5; SEQ ID NO 1301; 261pp + Sequence Listing; English.

CC The invention relates to identifying target proteins
 CC (ABB90790-ABB94016) for herbicidally active compounds, comprising
 CC aligning and comparing nucleic acid or amino acid sequences from plant
 CC with nucleic acid or amino acid sequences from non-plant organisms using
 CC suitable search parameters, where plant sequences having an E-value
 CC greater by a factor of 3 than the E-value of most similar non-plant
 CC sequences are selected. The polypeptides or nucleic acids encoding them
 CC are useful for identifying modulators. The identified modulators are
 CC useful as herbicides.

SQ Sequence 256 AA;

Query Match 49.5%; Score 627; DB 23; Length 256;
 Best Local Similarity 54.2%; Pred. NO. 5.9e-53;
 Matches 136; Conservative 38; Mismatches 65; Indels 12; Gaps 6;

QY 4 ALSGTAIVSTAALLAPIRAPTSAFIRRSOLTCRHLSTKCRAG-----SIYPAANAAGS 59
 Db 11 ALPCSSTKSSSLATFRSP---FLRFGNSTSLPSSISITRGRGSSPTIIPRAAASDS 67
 QY 60 SSP-SSAVFHGECFVVGDNIDTDQIIPAEHLTVPSKPDYKRLGSFAFAGLPSAAYPTP 118

Db 68 NEALANTTFHGLCYVLNDNDITDQIIPGAGACTFPPSNOERDEIAAHALSGLD-FHKTR 126
 QY 119 FVARGESSRAITIVGANGCCGSSREHAPVALGAMARAIVAEGARIFFRNSVATGEV 178
 Db 127 FIEPGENSKSIITIGBENCGSSRSHAPVYCAGAAKAIVAESYRIFRNSVATGEV 186
 QY 179 YPLELTDVGAMKECKTDPVYTVDLANS--VFINTTSCKEYKLKPIFDAGVIEAGGIFAY 236
 Db 187 FPLE-SEVRVCECKTGDYVYIELSDSGGLTNHTTOKNKIKNSIGDAGVIDAGGIFAY 245
 QY 237 ARKTGMASKA 247
 Db 246 ARMMGMIPSLA 256

RESULT 13
ABB92092
ID ABB92092 standard; Protein; 256 AA.

AC ABB92092;

DT 31-MAY-2002 (first entry)

DE Herbicidally active polypeptide SEQ ID NO 1303.

KW herbicidal; plant; agriculture; herbicide.

OS *Arabidopsis thaliana*.

PN WO200210210-A2.

PD 07-FEB-2002.

PF 28-AUG-2001; 2001WO-EP09892.

PR 28-AUG-2001; 2001WO-EP09892.

PA (FARB) BAYER AG.

PI Tietjen K, Weidler M;

DR WPI; 2002-269010/31.

PT Identifying plant target proteins for herbicidally active compounds

PT from plant with nucleic acid or amino acid sequences from non-plant

XX
CS
Classification: SEO ID NO 1303. 261nn + Sequence Listing: English

The function relates to identifying target proteins, (AB930790-AB94016) for herbicidally active compounds, comprising aligning and comparing nucleic acid or amino acid sequences from plant with nucleic acid or amino acid sequences from non-plant organisms using suitable search parameters, where plant sequences having an E-value greater by a factor of 5 than the E-value of nucleic acids encoding them sequences are selected. The polypeptides or nucleic acids encoding them are useful for identifying modulators. The identified modulators are useful as herbicides.

SQ Sequence 256 AA;

Query Match	49.58;	Score 627;	DB 23;	Length 256;
Best Local Similarity	54.28;	Pred. No. 5.9e-55;		
Matches 136;	Conservative 38;	Mismatches 65;	Indels 12;	Gaps 6

```

0Y      4  AATGAATVSTAALLAPIRAFTSAFIRRSQITCRHLSLCKRAG-----SYVAAAAAAGS  59
      || :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db      11  ALPCSTKTSISLATFRSP---FLRNGSTSLPSSISITSRGTSPTIIPRAASSDS  67
      || :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
0Y      60  SSP--SSAVHGECEFGVGDINDIDQIIIPAHLTLVSKDEYRKLGSFPAALPSAAAPTP  118
      || :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::

```

[illegible]

RESULT 14
AAB96359
ID AAB96359 standard; Protein; 164 AA

AAC 96359;

DT 29-OCT-2001 (first entry)

DE Putative 3-isopropylmalate dehydratase/aconitase small subunit #2.

KW Hyperthermophilic archaeon; hyperthermophilic protein.

05 Pyrococcus abyssi

PN FR2792651-A1.

PD 27-OCT-2000.

PF 21-APR-1999; 99FR-0005034.

PR 21-APR-1999; 99ER-0005034.

PA (CNRS) CNRS CENT NAT RECH SCI.
(TITE-) TERREMER INST EP RECH EXPL. MER

XX
XX
XXXXXX D. Bhierry TC Briaur D. Dietrich J. Lecompte O:

PI Querellou J, Weissenbach J, Saurin W, Hellig K

DR WPI; 2001-126236/14.

PT New nucleotide sequences isolated from *Pyrococcus abyssi* encode proteins useful in industry -

XX
DC
CJatw 7: Page 1040: 1657nd: French

The present invention relates to the genomic sequence of *Pyrococcus* abyssal (see AAH6431 and AAH1223-7) and *P. abyssal* proteins, *P. abyssal* is a hyperthermophilic archaeon, which is isolated from deep-sea hydrothermal vents. The present sequence is one such *P. abyssal* protein. The proteins of the present invention have various potential industrial uses, since the proteins are stable at very high temperatures, some up to 110 degrees centigrade.

Note: This patent is in the same patent family as WO200605062, which contains additional sequences as shown in AAB99132-AAB99143, AAH75903-AAH75920 and AAG66436.

Sequence	164 AA
5Q	

Query Match Similarity	20.43;	Score 258.5;	DB 22;	Length 164;
Best Local Similarity	39.08;	Pred. No. 5.5e-18;		
Matches 67; Conservative	26;	Mismatches 62;	Indels 17;	Gaps 7;

```

QY 69 GECFVVGNDIDQIIPAHHLTVSPKDEYKNCESFAPALPSPAATPTPVPAAGESSR 128
    | : | | | | | | | | | | : | : | : |
Db 5 GRAMKGDNDIDVILIPAYTL--TSDPE--LAKHVLDDLPBFSNK--MKGD---- 53
    | : | | | | | | | | | | : | : | : | : | : |
QY 129 YAIIVGANGCCSREHFPALGAAGARAIVAEGYARIFRNSVATGEVYPIELTDVGA 188
    | | | | | | | | | | : | : | : | | | : | |

```


Db 54 --IIVAGENFGCGSSREHAPLAIKAAGVSCVIAKSFARIFRNAINIG--LPI-LEAPQA 108

Job time : 38 secs

Qy 189 WKCKTGVDVYVDLANSVFINHTSGKEYKLPID-AGPVIAGGIFAYARK 239

Db 109 VDRIFTGDELEVDFSSGSEIRNLTKEGYVRANFPDFIMEITKAGGIVEMAKR 160

RESULT 15

AAB96132

ID AAB96132 standard; Protein; 163 AA.

XX AAB96132;

XX AC AAB96132;

XX DT 29-OCT-2001 (first entry)

XX DE Putative 3-isopropylmalate dehydratase/acnitase large subunit #2.

XX KW Hyperthermophilic archaeon; hyperthermophilic protein.

XX OS Pyrococcus abyssi.

XX PN FR2792651-A1.

XX PD 27-OCT-2000.

XX PE 21-APR-1999; 99PR-0005034.

XX PR 21-APR-1999; 99PR-0005034.

XX PA (CNRS) CNRS CENT NAT RECH SCI.

XX PI (IFRE-) IFREMER INST FR RECH EXPL MER.

XX PI Forterre P, Thierry JC, Prieur D, Dietrich J, Lecompte O;

XX PI Querellou J, Welssenbach J, Saurin W, Heilig R;

XX DR WPI; 2001-126236/14.

XX PT New nucleotide sequences isolated from Pyrococcus abyssi encode

XX PS Claim 7: Pages 768-769; 1657pp; French.

XX CC The present invention relates to the genomic sequence of Pyrococcus

XX CC abyssi (see AA86431 and AA4123-7) and P. abyssi proteins. P. abyssi is

XX CC a hyperthermophilic archaeon, which is isolated from deep-sea

XX CC hydrothermal vents. The present sequence is one such P. abyssi protein.

XX CC The proteins of the present invention have various potential industrial

XX CC uses, since the proteins are stable at very high temperatures, some up to

XX CC 110 degrees centigrade.

XX CC Note: This patent is in the same patent family as WO200065062, which

XX CC contains additional sequences as shown in AAB9132-AAB9143,

XX CC AAH75903-AAH75920 and AA66436.

XX SQ Sequence 163 AA;

XX Query Match 16.3%; Score 206.5; DB 22; Length 163;

XX Best Local Similarity 34.1%; Pred. No. 9.9e-13;

XX Matches 60; Conservative 29; Mismatches 62; Indels 25; Gaps 8;

Search completed: March 31, 2003, 09:07:38

GenCore version 5.1.4.P5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 31, 2003, 09:07:03 ; Search time 15 Seconds
(without alignments)
488.421 Million cell updates/sec

Title: US-10-027-450-47
Perfect score: 1266
Sequence: 1 MAALSGTAVSTAALLAPIR.....AGGIFAYARKTGIMASKAA 249

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: Issued_Patents_AA:*
2: /cgn2_6/ptodata/1/1aa/5A.COMB.pep:*
3: /cgn2_6/ptodata/1/1aa/5B.COMB.pep:*
4: /cgn2_6/ptodata/1/1aa/6A.COMB.pep:*
5: /cgn2_6/ptodata/1/1aa/6B.COMB.pep:*
6: /cgn2_6/ptodata/1/1aa/Backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1266	100.0	249	4	US-09-173-300-47
2	1226	96.8	244	4	US-09-173-300-53
3	994.5	78.6	257	4	US-09-173-300-49
4	740.5	58.5	263	4	US-09-173-300-51
5	191.5	15.1	191	1	US-08-403-866-4
6	166	13.1	189	4	US-09-134-001C-3886
7	115.5	9.1	189	2	US-08-887-798-2
8	98.5	7.8	914	4	US-09-134-001C-5208
9	88	7.0	461	4	US-09-355-115-7
10	87	6.9	119	4	US-08-858-207A-388
11	85.5	6.8	4550	2	US-08-804-227C-8
12	85.5	6.8	4550	2	US-08-804-198-2
13	83	6.6	742	4	US-08-791-115B-5
14	81	6.4	1544	4	US-09-413-814-6
15	80.5	6.4	1234	4	US-08-317-310A-15
16	80.5	6.4	1334	5	PCR-US95-130A1-15
17	79.5	6.3	378	4	US-09-325-932A-158
18	79.5	6.3	3739	3	US-09-320-878-2
19	79	6.2	1864	2	US-08-804-227C-3
20	78	6.2	514	4	US-09-385-028-5
21	78	6.2	622	4	US-09-134-001C-3882
22	78	6.2	731	2	US-08-911-364-1
23	78	6.2	733	4	US-08-464-700-2
24	78	6.2	792	4	US-08-678-039A-40
25	78	6.2	1890	4	US-09-004-838-88
26	77.5	6.1	1243	2	US-08-557-139-2
27	77.5	6.1	4472	2	US-08-804-227C-2

28	77.5	6.1	4545	2	US-08-804-227C-14	Sequence 14, Appl
29	76.5	6.0	430	2	US-08-809-740A-2	Sequence 2, Appl
30	76	6.0	387	4	US-09-323-427-4	Sequence 4, Appl
31	76	6.0	387	4	US-09-812-642-4	Sequence 4, Appl
32	76	6.0	970	1	US-08-375-709-7	Sequence 7, Appl
33	76	6.0	970	1	US-08-752-929-7	Sequence 7, Appl
34	76	6.0	970	4	US-09-090-793-5	Sequence 5, Appl
35	76	6.0	1261	1	US-08-764-100-26	Sequence 26, Appl
36	76	6.0	1854	4	US-09-004-838-108	Sequence 108, App
37	75.5	6.0	591	3	US-08-965-903B-2	Sequence 2, Appl
38	75.5	6.0	1298	2	US-08-690-473-2	Sequence 2, Appl
39	75.5	6.0	1298	4	US-09-259-821A-2	Sequence 2, Appl
40	75.5	6.0	1298	4	US-08-843-659-2	Sequence 2, Appl
41	75	5.9	505	1	US-08-221-750A-5	Sequence 5, Appl
42	75	5.9	3165	2	US-08-459-146-3	Sequence 3, Appl
43	75	5.9	3165	2	US-08-459-065-3	Sequence 3, Appl
44	75	5.9	3739	4	US-09-105-537-33	Sequence 33, Appl
45	75	5.9	11877	4	US-09-105-537-6	Sequence 6, Appl

ALIGNMENTS

```
RESULT 1
US-09-173-300-47
; Sequence 47, Application US/09173300
; Patent No. 6451581
; GENERAL INFORMATION:
; APPLICANT: Falco, Saverio Carl
; APPLICANT: Hiltz, William D.
; APPLICANT: Kinney, Anthony J.
; APPLICANT: Cahoon, Rebecca E.
; APPLICANT: Rafalski, J. Antoni
; TITLE OF INVENTION: PLANT BRANCHED CHAIN AMINO ACID BIOSYNTHETIC ENZYMES
; FILE REFERENCE: BB-1126
; CURRENT APPLICATION NUMBER: US/09/173, 300
; EARLIER FILING DATE: 1998-10-15
; EARLIER APPLICATION NUMBER: 60/063, 423
; EARLIER FILING DATE: 1997 October 28
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: Microsoft Word Version 7.0A
; SEQ ID NO 47
; LENGTH: 249
; TYPE: PRT
; ORGANISM: Zea mays
US-09-173-300-47

Query Match      100.0%  Score 1266; DB 4; Length 249;
Best Local Similarity 100.0%; Pred. No. 4.3e-138;
Matches 249; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAALSGTAVSTAALLAPIRAPTSAFIRRSQITCRHLSTLCRRAGSIVPAAAAAGSS 60
    |||||||
    1 MAALSGTAVSTAALLAPIRAPTSAFIRRSQITCRHLSTLCRRAGSIVPAAAAAGSS 60
DB 1 MAALSGTAVSTAALLAPIRAPTSAFIRRSQITCRHLSTLCRRAGSIVPAAAAAGSS 60

QY 61 SPSSAFHEHCPCVGNIDTDOITFAEHRTVPSKPDYRKLGSFAFGLPSAAYPTPV 120
    |||||||
    61 SPSSAFHEHCPCVGNIDTDOITFAEHRTVPSKPDYRKLGSFAFGLPSAAYPTPV 120
DB 61 SPSSAFHEHCPCVGNIDTDOITFAEHRTVPSKPDYRKLGSFAFGLPSAAYPTPV 120

QY 121 APGEESRYAIIIGVANGFCGSSREHAPVALCAAGARAIVAEGYARIFFRNSVATGEY 180
    |||||||
    121 APGEESRYAIIIGVANGFCGSSREHAPVALCAAGARAIVAEGYARIFFRNSVATGEY 180
DB 121 APGEESRYAIIIGVANGFCGSSREHAPVALCAAGARAIVAEGYARIFFRNSVATGEY 180

QY 181 LETLDVGAMKECKTGTVVDLANSVFNIHNSGKRYKPIGDPVTEAGIFAAYARKT 240
    |||||||
    181 LETLDVGAMKECKTGTVVDLANSVFNIHNSGKRYKPIGDPVTEAGIFAAYARKT 240
DB 181 LETLDVGAMKECKTGTVVDLANSVFNIHNSGKRYKPIGDPVTEAGIFAAYARKT 240

QY 241 GMTASKAAA 249
    |||||||
    241 GMTASKAAA 249
DB 241 GMTASKAAA 249

RESULT 2
```

```

US-09-173-300-53
; Sequence 53, Application US/09173300
; Patent No. 6451581
; GENERAL INFORMATION:
; APPLICANT: Falco, Saverio Carl
; APPLICANT: Hitz, William D.
; APPLICANT: Kinney, Anthony J.
; APPLICANT: Cahoon, Rebecca E.
; APPLICANT: Ratajski, J. Antoni
; TITLE OF INVENTION: PLANT BRANCHED CHAIN AMINO ACID BIOSYNTHETIC ENZYMES
; FILE REFERENCE: BB-1126
; CURRENT APPLICATION NUMBER: US/09/173,300
; CURRENT FILING DATE: 1998-10-15
; EARLIER APPLICATION NUMBER: 60/063,423
; EARLIER FILING DATE: 1997 October 28
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: Microsoft Word Version 7.0A
; SEQ ID NO 53
; LENGTH: 244
; TYPE: PRT
; ORGANISM: Triticum aestivum
US-09-173-300-53

```

Query Match	96.8%	Score 1226	DB 4	Length 244
Best Local Similarity	99.6%	Pred. No. 1.8e-133		
Matches 240: Conservative	0	Mismatches 1	Indels 0	Gaps 0

[illegible]

RESULT 3
US-09-173-300-49
; Sequence 49, Application US/09173300
; Patent No. 6451581

```

? APPLICANT: Falco, Saverio Carl
? APPLICANT: Hiltz, William D.
? APPLICANT: Kinney, Anthony J.
? APPLICANT: Cahoon, Rebecca E.
? APPLICANT: Rafalski, J Antoni
? TITLE OF INVENTION: PLANT BRANCHED CHAIN AMINO ACID BIOSYNTHETIC
? FILE REFERENCE: BB-1126
? CURRENT APPLICATION NUMBER: US/09/173,300
? CURRENT FILING DATE: 1998-10-15
? EARLIER APPLICATION NUMBER: 60/063,423
? EARLIER FILING DATE: 1997 October 28
? NUMBER OF SEQ ID NOS: 54
? SOFTWARE: Microsoft Word Version 7.0A
? SEQ ID NO 49
? LENGTH: 257
? TYPE: PRF
? ORGANISM: Oryza sativa
? US-09-173-300-49

```

Query Match	78.68;	Score 994.5;	DB 4;	Length 257;
-------------	--------	--------------	-------	-------------

	Best Local Similarity	77.0%: Pred. No. 1,1e-106;	Matches 196;	Conservative 18;	Mismatches 32;	Indels 9;	Gaps 2.
QY	1	MAALSGTAVASTA-----	ALLAPIRAPTSAEIRRSQLTCHRLSLKC---	RRAGSTVPA	51		
		: - -		- -	- -		
Db	1	MAAAAAAPALSLAEAPVTAIVLAPCPTPTSPRTFRRSWVAALCRALCKHRSREPLTAAVAA	60				
QY	52	AAAAAAGSSPSSASVPHGECFVYVGGDMDIDDOIIPAEHLITVPSKPDEKRGSAFAFGP	111				
Db	61	AAAAAAGDSTSGVHGECEVYVGGDMDIDDOIIPAEHLITVPSKPDEKRGISAFVGLP	120				
QY	112	SAAYPTFEVAPGESSRYATIVGAGNFGCGSSREHAPVALGAAGARAIVAEGVARIFFRN	171				
Db	121	TAAFTPTFFVAPGEGTTTAYIIIGGAFNFGGSSREHAPVALGAAGARAVAEVGARIFFRN	180				
QY	172	SVATGEVYVPELTDVGMKECKTGDDVYVYDVLANSVFINHTSKSEKYLKPIGDADAPVLEAG	231				
Db	181	SVATGEVYVPELTDVGMKECKTGDDVYVYDVLNLCVMIHNTSKQYKIKPIGDADAPVLEAG	240				
QY	232	GIFAVYARTKGMISKAA	248				
Db	241	GIFAVARTKGMISKAA	257				

RESULT 4
US-09-173-300-51
Sequence 51, Application US/09173300

```

? Patent No.6451581
? -----
? GENENT INFORMATION:
? APPLICANT: Falco, Saverio Carl
? APPLICANT: Hiltz, William D.
? APPLICANT: Kinney, Anthony J.
? APPLICANT: Cahoon, Rebecca E.
? APPLICANT: Rafalski, J. Antoni
? TITLE OF INVENTION: PLANT BRANCHED CHAIN AMINO ACID BIOSYNTHETIC
? FILE REFERENCE: BB-1126
? CURRENT APPLICATION NUMBER: US/09/173,500
? CURRENT FILING DATE: 1998-10-15
? EARLIER APPLICATION NUMBER: 60/063,423
? EARLIER FILING DATE: 1997 October 28
? NUMBER OF SEQ ID NOS: 54
? SOFTWARE: Microsoft word Version 7.0A
? SEQ ID NO 51
? LENGTH: 263
? TYPE: prt
? ORGANISM: Glycine max
? FEATURE:
? NAME/KEY: UNSURE
? LOCATION: (4)
?
US-09-173-300-51

```

Query Match	58.58;	Score 740.5;	DB 4;	Length 263;
Best Local Similarity	62.88;	Pred No 2.5e-77;		
Matches 157; Conservative	24;	Mismatches 52;	Indels 17;	Gaps 5;

```

QY      8  TAAVTAALLIAPRAPTAFAFIRKSQLTCHRL-----HSLKCRRAAGSIYPMAAAAA 56
Db     20  TRFSSAATLP--RNIAATTKLISHSTHLLPRELSFPTPKSSNPNNRVAVSLQTPAQS 76
QY     57  AGSSSPSAAVFHQEGEVVDNDITDQIIPAEHLTVPSKPDYEYKRLSGFAPGLSPSAYP 116
Db     77  AASASP-SASFHCLCYVGDNDITDQIIPAEYLLTVPSKPDYEYKRLSLAYLGLP-ATYA 134
QY     117  TPVPVAPGEESRRAIIVGANGFCSSSEHAPVALGAGARAIVAEGYARLFFENSVATG 176
Db     135  TRTRIEGEIKTKAIIVIGGANGFCSSSRHAPVALGAGAAVAEASARYLFFENSVATG 184
QY     177  EYVPLELITDYGAMKEKTDVYTVLPLANSVFNNHSGKEYLKLPGDGPVYTEAGGFAY 236
Db     195  EYVPLE-SGEKCEBCTGDDVYVIELGESRLNHTTGKEYRLKPLGDGAPVITEAGGIFAY 255
QY     237  ARRTGMIAK 246
      |||||||:

```

Db 254 ARKGMIPSR 263

RESULT 5

US-08-403-866-4

Sequence 4, Application US/08403866

Patent No. 5643779

GENERAL INFORMATION:

APPLICANT: Ehrlich, Stanislaw

APPLICANT: Godon, Jean-Jacques

APPLICANT: Renault, Pierre

TITLE OF INVENTION: Nucleic acid coding for an alpha-acetolactate

TITLE OF INVENTION: synthase from Lactococcus and its applications

NUMBER OF SEQUENCES: 16

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/403,866

FILING DATE:

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Goldman, Michael L.

REGISTRATION NUMBER: 30, 727

REFERENCE/DOCKET NUMBER: 20747/30

TELECOMMUNICATION INFORMATION:

TELEPHONE: (716) 263-1487

TELEFAX: (716) 263-1487

TELEX: 978450 (WUT)

INFORMATION FOR SEQ ID NO: 4:

SEQUENCE CHARACTERISTICS:

LENGTH: 191 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

HYPOTHEICAL: NO

ORIGINAL SOURCE:

ORGANISM: Lactococcus lactis subsp. lactis

INDIVIDUAL ISOLATE: LEUD

US-08-403-866-4

Query Match

Best Local Similarity 15.1%; Score 191.5; DB 1; Length 191;

Matches 58; Conservative 35; Mismatches 70; Indels 37; Gaps 7;

QY 58 GSSPSAVHGCFFVGVNDIDDIIPAEHLTVPSKDEYRKLGSFAPLPSAAPT 117

DB 9 GTSTP-----VANDIDIDDIIPAEHLTVPSKDEYRKLGSFAPLPSAAPT 56

QY 118 PVAAGESSRYAIVYGANGCGSSREHAPVALGAGARAIYAGYARIFFRNSVATGE 177

DB 57 DFILNARKYKASILLISGDFGSSREHAPVALGAGARAIYAGYARIFFRNSVATGE 115

QY 178 VYPL-----ELTDVGAKKEKTDGVTVVDLANSVFINTSGKEYKIKPL----- 221

DB 116 LKPKOPREVLNIDTKISSOE-----TIDLPQLITSLGDFHFEIDPIWKDLING 169

QY 222 -GDAGPYIE-AGGIFAYAR 239

DB 170 LDDIGITLQYEAISAVEOK 189

RESULT 6

US-09-134-001C-3886

Sequence 3886, Application US/09134001C

Patent No. 6380370

GENERAL INFORMATION:

APPLICANT: Lynn Doucette-Stamm et al

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS

TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS

FILE REFERENCE: GTC-007

Query Match

Best Local Similarity 9.1%; Score 115.5; DB 2; Length 780;

Matches 52; Conservative 11; Mismatches 56; Indels 16; Gaps 5;

QY 76 DNIDTQIIPAEHLTVPSKDEYRKLGSFAP---AGLPSAAPTYPVAPGESSRYAIT 132

DB 19 DNIDTQIIPAEHLTVPSKDEYRKLGSFAP---AGLPSAAPTYPVAPGESSRYAIT 73

QY 133 VGGANGCGSSREHAPVALGAGARAIYAGYARIFFRNSVATGEVYPLELTDVGAMKE- 191

DB 74 ITGDNFGCGSSREHAPVALGAGARAIYAGYARIFFRNSVATGEVYPLELTDVGAMKE- 128

QY 192 --CKTGDVTVVDLAN 204

DB 129 EHLAOFDEITVDLPN 143

RESULT 7

US-08-887-798-2

Sequence 2, Application US/08887798

Patent No. 5922536

GENERAL INFORMATION:

APPLICANT: Mayeux, Richard

APPLICANT: Graziano, Joseph H.

TITLE OF INVENTION: PARKINSON'S DISEASE TESTS

NUMBER OF SEQUENCES: 38

CORRESPONDENCE ADDRESS:

ADDRESS: Cooper & Dunham LLP

STREET: 1185 Avenue of the Americas

CITY: New York

STATE: New York

COUNTRY: U.S.A.

ZIP: 10036

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/887,798

FILING DATE:

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: White, John P.

REGISTRATION NUMBER: 28, 678

REFERENCE/DOCKET NUMBER: 0575/51949/JPW

TELECOMMUNICATION INFORMATION:

TELEPHONE: 212-278-0400

TELEFAX: 212-391-0525

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 780 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-887-798-2

Query Match

Best Local Similarity 23.3%; Score 115.5; DB 2; Length 780;

Matches 52; Conservative 11; Mismatches 56; Indels 16; Gaps 5;

QY 76 DNIDTQIIPAEHLTVPSKDEYRKLGSFAP---AGLPSAAPTYPVAPGESSRYAIT 132

DB 19 DNIDTQIIPAEHLTVPSKDEYRKLGSFAP---AGLPSAAPTYPVAPGESSRYAIT 73

QY 133 VGGANGCGSSREHAPVALGAGARAIYAGYARIFFRNSVATGEVYPLELTDVGAMKE- 191

DB 74 ITGDNFGCGSSREHAPVALGAGARAIYAGYARIFFRNSVATGEVYPLELTDVGAMKE- 128

QY 192 --CKTGDVTVVDLAN 204

DB 129 EHLAOFDEITVDLPN 143

121 -APGESSRYAILVGGANFGCCSSR-----EHAP-----VALGAA-----GARAI 155

GenCore version 5.1.4.p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 31, 2003, 09:08:58 ; Search time 15 Seconds
(without alignments)
974.813 Million cell updates/sec

974.813 Million cell updates/sec

Title: US-10-027-450-47
Perfect score: 1266
Sequence: 1 MAALSGTAVSTALLAPIR.....ACGIFAVARKTGIMASKAA 249

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 237916 seqs, 58723674 residues

Total number of hits satisfying chosen parameters: 237916

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published_Applications_AA:*
1: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
2: /cgn2_6/ptodata/2/pubpaa/PC1_NEW_PUB.pep:*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
7: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*
10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep:*
11: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*
12: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep:*
13: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*
14: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1266	100.0	249	US-10-027-450-47	Sequence 47, Appl
2	1226	96.8	244	US-10-027-450-53	Sequence 53, Appl
3	994.5	78.6	257	US-10-027-450-49	Sequence 49, Appl
4	740.5	58.5	263	US-10-027-450-51	Sequence 51, Appl
5	184	14.5	212	US-09-815-242-11900	Sequence 11900, A
6	164.5	13.0	644	US-09-815-242-5730	Sequence 5730, Ap
7	155.5	12.3	201	US-09-815-242-14016	Sequence 14016, A
8	155	12.2	190	US-09-815-242-12157	Sequence 12157, A
9	153.5	12.1	201	US-09-815-242-10020	Sequence 10020, A
10	151	11.9	200	US-09-815-242-11165	Sequence 11165, A
11	134	10.6	197	US-09-738-626-4953	Sequence 4953, Ap
12	130	10.3	943	US-09-738-626-5199	Sequence 5199, Ap
13	120	9.5	200	US-09-815-242-11648	Sequence 11648, A
14	112	8.8	417	US-09-925-302-700	Sequence 700, Appl
15	105.5	8.3	538	US-09-779-307-2	Sequence 2, Appl
16	105.5	8.3	538	US-09-779-307-11	Sequence 11, Appl
17	101	8.0	547	US-09-779-307-12	Sequence 12, Appl
18	101	8.0	547	US-09-779-307-13	Sequence 13, Appl
19	98	7.7	963	US-09-924-396B-20	Sequence 20, Appl

20	94.5	7.5	1051	9	US-09-976-059-13	Sequence 13, Appl
21	93	7.3	952	9	US-09-924-396B-18	Sequence 18, Appl
22	90	7.1	444	10	US-09-864-761-42633	Sequence 42633, A
23	88.5	7.0	341	9	US-09-738-626-4543	Sequence 4543, Ap
24	88.5	7.0	585	9	US-09-738-626-6032	Sequence 6032, Ap
25	84.5	6.7	324	9	US-09-738-626-4887	Sequence 4887, Ap
26	84	6.6	354	10	US-09-825-414-70	Sequence 70, Appl
27	83	6.6	537	9	US-09-738-626-4264	Sequence 4264, Ap
28	82.5	6.5	1056	9	US-10-161-510-10	Sequence 10, Appl
29	81	6.4	442	10	US-09-749-728B-11	Sequence 11, Appl
30	81	6.4	562	10	US-09-815-242-12159	Sequence 12159, A
31	81	6.4	562	10	US-09-815-242-12929	Sequence 12929, A
32	80.5	6.4	567	9	US-09-344-882-30	Sequence 30, Appl
33	80	6.3	351	10	US-09-815-242-11984	Sequence 11984, A
34	79	6.2	345	10	US-09-789-836-5	Sequence 5, Appl
35	79	6.2	430	9	US-10-029-180-32	Sequence 32, Appl
36	78	6.2	354	10	US-09-999-152-14	Sequence 14, Appl
37	77.5	6.1	398	9	US-09-738-626-6532	Sequence 6532, Ap
38	77.5	6.1	460	9	US-09-976-673-16	Sequence 16, Appl
39	77.5	6.1	470	9	US-09-976-673-18	Sequence 18, Appl
40	77.5	6.1	1242	9	US-09-436-184-5	Sequence 5, Appl
41	77.5	6.1	1242	10	US-09-903-248-5	Sequence 5, Appl
42	77.5	6.1	1242	10	US-09-859-604-5	Sequence 5, Appl
43	77.5	6.1	1242	10	US-09-903-063-5	Sequence 5, Appl
44	77.5	6.1	1242	10	US-09-903-216-5	Sequence 5, Appl
45	77.5	6.1	1242	10	US-09-903-199-5	Sequence 5, Appl

ALIGNMENTS

RESULT 1
US-10-027-450-47
Sequence 47, Application US/10027450
Patent No. US20020102715A1
GENERAL INFORMATION:
APPLICANT: Falco, Saverio Carl
APPLICANT: Hitz, William D.
APPLICANT: Kinney, Anthony J.
APPLICANT: Cahoon, Rebecca E.
APPLICANT: Rafalski, J. Antoni
TITLE OF INVENTION: PLANT BRANCHED CHAIN AMINO ACID BIOSYNTHETIC ENZYMES
FILE REFERENCE: BB-1126
CURRENT APPLICATION NUMBER: US/10/027, 450
PRIOR APPLICATION NUMBER: 2001-12-20
PRIOR FILING DATE: 1997 October 28
NUMBER OF SEQ ID NOS: 54
SOFTWARE: Microsoft Word Version 7.0A
SEQ ID NO 47
LENGTH: 249
TYPE: PRT
ORGANISM: Zea mays
US-10-027-450-47

Query Match 100.0%; Score 1266; DB 12; Length 249;
Best Local Similarity 100.0%; Pred. No. 1.2e-113;
Matches 249; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 MAALSGTAVSTALLAPIRAPTSAFIRRSQUTCHRLSLKCRAGSIVPAAAAAGSS 60
DB 1 MAALSGTAVSTALLAPIRAPTSAFIRRSQUTCHRLSLKCRAGSIVPAAAAAGSS 60
OY 61 SPSSAVFHECEVVDNDITDIIIPAEHLITVPSKPDYRKLGSFAAGLPSAAYPPPV 120
DB 61 SPSSAVFHECEVVDNDITDIIIPAEHLITVPSKPDYRKLGSFAAGLPSAAYPPPV 120
OY 121 APEESRFAIIVGANGCGSSREHAPVALCAAGARAIIVEGARIFFRNSVATGYVP 180
DB 121 APEESRFAIIVGANGCGSSREHAPVALCAAGARAIIVEGARIFFRNSVATGYVP 180
OY 181 LELVDGAMKECKTDVVTDLANSVFTINHTSGKEYKIKPIGDAGPVIEAGGIFAVARKT 240
DB 181 LELVDGAMKECKTDVVTDLANSVFTINHTSGKEYKIKPIGDAGPVIEAGGIFAVARKT 240

Dd	181	LELDYDVGMMKCKCGDYYTVVDLANSVFLNHISGKNEYLKPIGDAGPIEAGCIFATYARKT	240
Qy	241	GMIASKAAA	249
Dd	241	GMIASKAAA	249

RESULT 2
US-10-027-450-53
; Sequence 53, Application US/10027450
; Patent No. US20020102715A1

```

: APPLICANT: Falco, Saverio Carl
: APPLICANT: Hitz, William D.
: APPLICANT: Kinney, Anthony J.
: APPLICANT: Cahoon, Rebecca E.
: APPLICANT: Rafalski, J. Antoni
: TITLE OF INVENTION: PLANT BRANCHED CHAIN AMINO ACID BIOSYNTHETIC ENZYMES
: FILE REFERENCE: BB-1126
: CURRENT APPLICATION NUMBER: US/10/027,450
: CURRENT FILING DATE: 2001-12-20
: PRIOR APPLICATION NUMBER: 60/063,423
: PRIOR FILING DATE: 1997 October 28
: NUMBER OF SEQ. ID NOS: 54
: SOFTWARE: Microsoft Word Version 7.0A
: SEQ. ID NO. 53
: LENGTH: 244
: TYPE: PRT
: ORGANISM: Trillium aestivum
: US-10-027-450-53

```

Query Match	96.8%;	Score 1226;	DB 12;	Length 244;	.
Best Local Similarity	99.6%;	Pred. No. 7.9e-110;			
Matches 240; Conservative	0;	Mismatches 1;	Indels 0;	Gaps 0;	

Qy	9	AVSTAAALAPIRAPPTSAVIRRSQJLTCHEHLKCRKRGSIYPAAPAAAAAGSSPSSSAVFH	68
Db	4	AVSTAAALAPIRAPPTSAVIRRSQJLTCHEHLKCRKRGSIYPAAPAAAAAGSSPSSSAVFH	63
Qy	69	GECEFYVGNIDTDOIIPAHEHLTLVPSKDEYRKSGSFAGLPSAAVPTFPVAFGEBSR	128
Db	64	GECEFYVGNIDTDOIIPAHEHLTLVPSKDEYRKSGSFAGLPSAAVPTFPVAFGEBSR	123
Qy	129	YAIIVGANGFEGCGSSREHAPALVLAAGRAIYVAGSYARIFPRNSVANGVEYPLELTGVGA	188
Db	124	YAIIVGANGFEGCGSSREHAPALVLAAGRAIYVAGSYARIFPRNSVANGVEYPLELTGVGA	183
Qy	189	WKECTGTVVTVDLANSYFINHTSGKEYKLKPIDAGPVIIEAGSIFAYARKTGMIASKAA	248
Db	184	WKECTGTVVTVDLANSYFINHTSGKEYKLKPIDAGPVIIEAGSIFAYARKTGMIASKAA	243
Qy	249	A 249	
Db	244	A 244	

```

RESULT 3
US-10-027-450-49
: Sequence 49, Application US/10027450
: Patent No. US20020102715A1
: GENERAL INFORMATION:
: APPLICANT: Falco, Saverio Carl
: APPLICANT: Hiltz, William D.
: APPLICANT: Kinney, Anthony J.
: APPLICANT: Cahoon, Rebecca E.
: APPLICANT: Ratslaski, J. Antoni
: TITLE OF INVENTION: PLANT BRANCHED CHAIN AMINO ACID BIOSYNTHETIC ENZYMES
: FILE REFERENCE: BB-1126
: CURRENT APPLICATION NUMBER: US/10/027,450
: CURRENT FILING DATE: 2001-12-20
: PRIOR APPLICATION NUMBER: 60/063,423
: PRIOR FILING DATE: 1997 October 28
: NUMBER OF SEQ ID NOS: 54

```

```

; SOFTWARE: Microsoft Word Version 7.0A
; SEQ ID NO 49
; LENGTH: 257
; TYPE: PR1
; ORGANISM: Oryza sativa
US-10-027-450-49

```

Query Match	78.68;	Score 994.5;	DB 12;	Length 257;
Best Local Similarity	77.08;	Pred. No. 1.3e-87;		
Matches 198; Conservative	18;	Mismatches 32;	Indels 9;	Gaps 2;

Oy	MAALISGAIVSTA-----ALLAPIRAPVSAFIRRSOLTCHRLILSLKC---RAGASITVA	51
Db	1 MAAAAAALSLAEAPVATYALACPTPSRRFRSSRWAAICRALCHHSRPILTAVAA	60
Oy	52 AAAAASSSSSVAVFHGECPVVDNIDTDOIIPAHLTLVPSPDERYKLGSPAFAGLP	111
Db	61 AAAAAAGDSTISAGVFHGECPVVDNIDTDOIIPAHLTLVPSPDERYKLGSPAFAGLP	120
Oy	112 SAAAPTTPVAVGESSRAIIVYGANCGGSRPHAVAIAGAARIVAVEGARIFERN	171
Db	121 TAAAPTTPVAAEGEETRIAYIIIGANCGGSSRFHAHYALAGAGARVAAEGARIFERN	180
Oy	172 SVATGEVYPDELTVGAMKECKTDVVYVDLANSVLTNNHTSGKEYKLKP1GDAGPYTEAG	231
Db	181 SVAAGEVYPPELATJTGAMKECKTDGVVTVELDNQVMINHTSGKOYKLKP1GDAGPYTEAG	240
Oy	232 GIFAVARRKTGMIAASKAA	248
Db	241 GIFAVARRKTGMIAASKSA	257

RESULT 4
US-10-027-450-51
; Sequence 51, Application US/10027450

```

: GENERAL INFORMATION:
: APPLICANT: Falco, Saverio Carl
: APPLICANT: Hitz, William D.
: APPLICANT: Kinney, Anthony J.
: APPLICANT: Cahoon, Rebecca E.
: APPLICANT: Rafalski, J. Antoni
: TITLE OF INVENTION: PLANT BRANCHED CHAIN AMINO ACID BIOSYNTHETIC ENZYMES
: FILE REFERENCE: BB-1126
: CURRENT APPLICATION NUMBER: US/10/027,450
: CURRENT FILING DATE: 2001-12-20
: PRIOR APPLICATION NUMBER: 60/063,423
: PRIOR FILING DATE: 1997 October 28
: NUMBER OF SEQ ID NOS: 54
: SOFTWARE: Microsoft Word Version 7.0A
: SEQ ID NO 51
: LENGTH: 263
: TYPE: PRT
: ORGANISM: glycine max
: FEATURE:
: NAME/KEY: UNSURE
: LOCATION: (4)
: US-10-027-450-51

```

Query Match	58.5%	Score 740.5	DB 12	Length 263
Best Local Similarity	62.8%	Pred. No. 2,8e-63		
Matches 157	Conservative 24	Mismatches 52	Indels 17	Gaps 5

OY	8	TAVSTAALLAPIRAPTSAFIRRSQLTCHRL-----HSLKCRAGSIYPAAAAAA	56
	1	1	1
	1	1	1
Db	20	TRESSATATVP---RNIAFTKLISHSHYLLPRLFPETPKSSPNRRNVAVSLQTPRAOS	76
	1	1	1
OY	57	AGSSSPAAVHGECEVYGGNDITDOIIPAEHLTVLPSPKPEYKRLSPFAGLPSAAY	116
	1	1	1
	1	1	1
Db	77	AAASNP-SAFHGCLTYVGGNDITDOIIPAEHLTVLPSPKPEYKRLSYALIGLP-ATYA	134
	1	1	1
OY	117	TPVAPGESSRYAIIYVGGANFGCCSSREHAPVAGAAGAAIYAEGYARIFPENSAYTG	176
	1	1	1
	1	1	1

Db 135 TREIEPEIKTYAIVIGANGFCSSREHAPVALGASGAAYVAESYARIFFRNSVATG 194
Qy 177 EYVPLEITVGAMKECKTDVYTVDLANSVFHNHSGKEVYKIPGDDGPIYIEGTFAY 236
Db 195 EYVPLE-SEBRLCECTGDDVYVIELGESRLNHTTGKEYRKLPGDGPVIEAGLIPAY 253
Qy 237 ARKTMASK 246
Db 254 ARKTMIPSR 263

RESULT 5

US-09-815-242-11900
; Sequence 11900, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Cair, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11900
; LENGTH: 212
; TYPE: PRF
; ORGANISM: Pseudomonas aeruginosa
US-09-815-242-11900

Query Match 14.5%; Score 184; DB 10; Length 212;
Best Local Similarity 32.1%; Pred. No. 4.2e-10;
Matches 53; Conservative 23; Mismatches 45; Indels 44; Gaps 7;

Qy 77 NIPDDOIIIPAEHLTV-----PSKDEYRKGSFAFGLPAAVPTFVAAGDESS--- 127
Db 18 NVDTDOIIPKQPLKSIKRGFGPNLFDEKRYL-----DVGOPGDNSKRP 62
Qy 128 -----RY---AIVGANFGCGSSREHAPVALGAAGARAIVAGYARIFFRNSVA 174
Db 63 LNPDEVLPNORYOGASVLLARENFVGCGSSREHAPVALDEYGFRTVIADSYADIEFFNSFK 122
Qy 175 TGEVYPL-----ELTDVGAMKECKTDGVTVDLANSVFHNHSGK 214
Db 123 NG-LIPILPEAVDELFRQVANEQYQSLDIAAQT-VTRPDGK 165

RESULT 6
US-09-815-242-5730
; Sequence 5730, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:

; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Cair, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5730
; LENGTH: 644
; TYPE: PRF
; ORGANISM: Staphylococcus aureus
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (1)...(644)
; OTHER INFORMATION: Xaa = Any Amino Acid
US-09-815-242-5730

Query Match 13.0%; Score 164.5; DB 10; Length 644;
Best Local Similarity 33.3%; Pred. No. 1.4e-07;
Matches 53; Conservative 12; Mismatches 51; Indels 43; Gaps 5;

Qy 41 KCRRAGSIVPAAAAAASSSSPSAVFEGCEV-----VIG 75
Db 427 KCARFHLVSPAMAAAA-----LHGK-FDYVRKVYVMAIKPILTYKGIYPLFN 476
Qy 76 DNIDDOIIPAEHLTVPSKDEYRKGSFAF---AGLPSAAYPTFVAAGDESSRYATI 132
Db 477 DNIDDOIIPKYLKRIKSKS-----GSGPFAFDEKRYLPDGSNDPDPNPKPKQIKASIL 531
Qy 133 VGANFGCGSSREHAPVALGAAGARAIVAGYARIFFRN 171
Db 532 ITGDNFGCGSSREHAAWALKDYGFIHIIAGSFDIFVYN 570

RESULT 7
US-09-815-242-14016
; Sequence 14016, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Cair, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242

Db 104 PSFADIFYGNSF--NNGLFVKLSD 126

RESULT 10
US-09-815-242-11165
Sequence 1165, Application US/09815242
Patent No. US20020061569A1
GENERAL INFORMATION:

APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Karl L.
APPLICANT: Zyskind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
APPLICANT: Xu, H. Howard
TITLE OF INVENTION: Identification of Essential Genes in
FILE REFERENCE: ELITRA.011A
CURRENT APPLICATION NUMBER: US/09/815,242
CURRENT FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 11165
LENGTH: 200
TYPE: PRT
ORGANISM: Haemophilus influenzae
US-09-815-242-11165

Query Match 11.9%; Score 151; DB 10; Length 200;
Best Local Similarity 28.2%; Pred. No. 5,6e-07;
Matches 48; Conservative 21; Mismatches 59; Indels 42; Gaps 6;

Db 45 AGSIVPAAAAAAGSSPSSAFHGECPYVGNITDQIIIPAEHLTLVPSKP-----DE 98
8 SGLVPLDAA-----NVDIDAIIPKQFLQATIRIGFGKHLFHE 45
QY 99 YKRLGSFAPAGLPSAAYPTP-FVAPGESSRYAIIIVGANFEGCGSSRHHAPVALGAGAR 157
46 WRYLD-----VEGTKPMEFEVLNPOYGATITLLARKNLGCGSSRHHAPWALADYGFK 98
QY 158 AIYABEGYARIFRNNSVATGEVYPL-----ELPDVGAKMECKTGDDVYVTDL 202
Db 99 VMTAPSFADIFYNNSL--NNHMLPIRLSEBEVEITQWWMANEGKQIHVDL 147

RESULT 11
US-09-738-626-4953
Sequence 4953, Application US/09738626
Publication No. US20020197605A1
GENERAL INFORMATION:

APPLICANT: NAKAGAWA, SATOSHI
APPLICANT: MIZOGUCHI, HIROSHI
APPLICANT: ANDO, SEIKO
APPLICANT: HAYASHI, MIKIRO
APPLICANT: OCHIAI, KEIKO
APPLICANT: YOKOI, HARUHIKO
APPLICANT: TATEISHI, NAOKO

APPLICANT: SENOH, AKIHIRO
APPLICANT: IKEDA, MASATO
APPLICANT: OZAKI, AKIO
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
FILE REFERENCE: 249-125
CURRENT APPLICATION NUMBER: US/09/738,626
CURRENT FILING DATE: 2000-12-18
PRIOR APPLICATION NUMBER: JP 99/377484
PRIOR FILING DATE: 1999-12-16
PRIOR APPLICATION NUMBER: JP 00/159162
PRIOR FILING DATE: 2000-04-07
PRIOR APPLICATION NUMBER: JP 00/280988
PRIOR FILING DATE: 2000-08-03
NUMBER OF SEQ ID NOS: 7059
SOFTWARE: PatentIn ver. 3.0
SEQ ID NO 4953
LENGTH: 197
TYPE: PRT
ORGANISM: Corynebacterium glutamicum
US-09-738-626-4953

Query Match 10.6%; Score 134; DB 9; Length 197;
Best Local Similarity 30.2%; Pred. No. 2,3e-05;
Matches 52; Conservative 20; Mismatches 62; Indels 38; Gaps 6;

QY 77 NIDTDQIIIPAEHLTLVPSKPDEYKLGSAFAPAGLPSAAYPT-----PVAPGESSRYA 130
Db 18 MNDTDQIIIPAVYKRYTR-----TGFEDGLFSWMRONDPNPLMDTYKNGS 64
QY 131 IIVGANFEGCGSSRHHAPVALGAGARAIIVAGYARIPFRNSVATGEVYPT-LELTDVG-A 188
Db 65 VLVAGPDFGTGSSRHHAPWALMDYGFRAVPSRFRADIFRSGAKAGLITGMEQSDLELL 124
QY 189 WK--ECKTGDVYVTDIANSVFTHNTSGKEYKIKPIGDAGPYIEAGGIFAVAR 238
Db 125 WKMEQTPGLELVNLE-----KQIVTAGDVVISFEVDYPTIR 161

RESULT 12
US-09-738-626-5199

Sequence 5199, Application US/09738626
Publication No. US20020197605A1
GENERAL INFORMATION:
APPLICANT: NAKAGAWA, SATOSHI
APPLICANT: MIZOGUCHI, HIROSHI
APPLICANT: ANDO, SEIKO
APPLICANT: HAYASHI, MIKIRO
APPLICANT: OCHIAI, KEIKO
APPLICANT: YOKOI, HARUHIKO
APPLICANT: TATEISHI, NAOKO
APPLICANT: SENOH, AKIHIRO
APPLICANT: IKEDA, MASATO
APPLICANT: OZAKI, AKIO
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
FILE REFERENCE: 249-125
CURRENT APPLICATION NUMBER: US/09/738,626
CURRENT FILING DATE: 2000-12-18
PRIOR APPLICATION NUMBER: JP 99/377484
PRIOR FILING DATE: 1999-12-16
PRIOR APPLICATION NUMBER: JP 00/159162
PRIOR FILING DATE: 2000-04-07
PRIOR APPLICATION NUMBER: JP 00/280988
PRIOR FILING DATE: 2000-08-03
NUMBER OF SEQ ID NOS: 7059
SOFTWARE: PatentIn ver. 3.0
SEQ ID NO 5199
LENGTH: 943
TYPE: PRT
ORGANISM: Corynebacterium glutamicum
US-09-738-626-5199

Query Match 10.3%; Score 130; DB 9; Length 943;
Best Local Similarity 23.1%; Pred. No. 0.00046;

	Matches	61;	Conservative	19;	Mismatches	97;	Indels	53;	Gaps	10;		
OY	17	APRAPTS	AIFRRSQTCHRLHSLKCR	RAGSI	VPAAAAAAGSSSPSSAVF	HGE	CT	VG	D	76		
		115	APALSPRSSASSADAK	SPSS-----	PRGSILLD	AGAGAGGAGG	SPCSN-----	RTSGI		165		
OY	77	NIDPD	LIIPAEHLTVPSK	PKDEYRKLGS	PAFAFGJPSAAV	PPYFAP	CESSRA	II	VG-	135		
		166	SMGIDQ----	RHSSPLPACP-----	CLFGPL	LACAPAGV	-SPG	GVSA	PEL	HAAL 211		
OY	136	-----	ANEGCGSSREHA	PVALCAAGARA	IVAE	GYARLR	FRRNSVAT	GEV	PLE	LT	DVG 187	
		212	DRLYAQRPA	FGCGE	SRHSY	PPALSP	ALAGAGV	GAAGP	LER	GA	PGR-----	HSVT 265
OY	188	AMKECKTG----	DVVT	VD	LANS	VFINH	SGREY	KLKP	IGD	AG	P	TEAGGI 233
		266	GYGDCAVGAR	YD	DEL	TALL	RLTY-----	GTGGGR-----	AGANG	EP	SGI 304	

Search completed: March 31, 2003, 09:13:33
Job time : 16 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Comugen Ltd.

OM protein - protein search, using sw model

Run on: March 31, 2003, 09:06:23 ; Search time 44 Seconds
(without alignments)

544.033 Million cell updates/sec

Title: US-10-027-450-47

Sequence: 166
1 MAALSGTAVSTALAPIR.....AGGIFAYARKGTMIASKKAA 249

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	710.5	56.1	251	2 H84861	3-isopropylmalate
2	651	51.4	253	2 T47781	3-isopropylmalate
3	627	49.5	256	2 A84862	3-isopropylmalate
4	311.5	24.6	166	2 A72363	3-isopropylmalate
5	272	21.5	161	2 E69328	3-isopropylmalate
6	271	21.4	219	2 E70421	3-isopropylmalate
7	267.5	21.1	168	2 D64459	3-isopropylmalate
8	266	21.0	170	2 E69210	3-isopropylmalate
9	260	20.5	163	2 B97290	3-isopropylmalate
10	258.5	20.4	164	2 E69051	3-isopropylmalate
11	252.5	19.9	162	2 E69051	3-isopropylmalate
12	250.5	19.8	170	2 E69458	3-isopropylmalate
13	249	19.7	165	2 H69469	3-isopropylmalate
14	229.5	18.1	208	2 F75353	3-isopropylmalate
15	227.5	18.0	166	2 D72394	3-isopropylmalate
16	223.5	17.7	165	2 H90418	3-isopropylmalate
17	206.5	16.3	163	2 A75160	3-isopropylmalate
18	202.5	16.0	163	2 G71180	3-isopropylmalate
19	192	15.2	659	2 F70453	3-isopropylmalate
20	191.5	15.1	191	2 E36889	3-isopropylmalate
21	189	14.9	202	2 A81983	3-isopropylmalate
22	188	14.8	215	2 G82364	3-isopropylmalate
23	184	14.5	212	2 F83255	3-isopropylmalate
24	184	14.5	689	2 S26864	3-isopropylmalate
25	183.5	14.5	758	2 T39210	3-isopropylmalate
26	178	14.1	199	2 C69650	3-isopropylmalate
27	174.5	13.8	193	2 A81323	3-isopropylmalate
28	174.5	13.8	194	2 G84031	3-isopropylmalate
29	172	13.6	200	2 S75839	3-isopropylmalate

30	171.5	13.5	202	2 B87273	3-isopropylmalate
31	171.5	13.5	208	2 I39571	probable 3-isoprop
32	170	13.4	213	2 E81128	3-isopropylmalate
33	169.5	13.4	193	2 AG1694	3-isopropylmalate
34	169	13.3	213	2 C81836	probable 3-isoprop
35	168	13.3	200	2 A82071	3-isopropylmalate
36	167.5	13.2	216	2 AE2919	3-isopropylmalate
37	167.5	13.2	177	2 F97693	3-isopropylmalate
38	161.5	12.8	177	2 H75373	3-isopropylmalate
39	159	12.6	190	2 B89998	3-isopropylmalate
40	158.5	12.5	211	2 AB3561	3-isopropylmalate
41	155.5	12.3	201	2 AC0516	3-isopropylmalate
42	155	12.2	201	2 S07306	3-isopropylmalate
43	155	12.2	779	2 S64011	3-isopropylmalate
44	154.5	12.2	201	2 C90638	3-isopropylmalate
45	154.5	12.2	201	2 C85489	isopropylmalate is

ALIGNMENTS

RESULT 1

H84861
3-isopropylmalate dehydratase, small subunit [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C:Accession: H84861
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.T.; Town, C.D.; Fujii, C.
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter
Nature 402, 761-768, 1999
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A:Reference number: A84420; MUID:20083487; PMID:10617197
A:Accession: H84861
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-251 <STO>
A:Cross-references: GB:AB002093; NID:93763918; PIDN:AAC64298.1; GSPDB:GN00139*
C:Genetics:
A:Gene: At2g43090
A:Map position: 2

Query Match 56.1%; Score 710.5; DB 2; Length 251;
Best Local Similarity 59.5%; Pred. No. 4e-51;
Matches 153; Conservative 27; Mismatches 62; Indels 15; Gaps 6;

QY 1 MAALSGTAVSTALAPIRAPTSFIRSQLTCHRLSKCRAGSIVPAAAAAGSS 60
DB 1 MAALSGANPTLSRLTASPNKPSFATRSFPL--RMSIV--ASNFKPLVSRHAASSF 56
QY 61 SPSSA-----VHGECPVGDNDITDQITPAEHLTVPSKPDYRKLGSAFAGLPSA 113
DB 57 VTRSAEPQERKTFHGLTYVGDNDITDQIIPAEFLTVPSNPPEYKLTGYALVGLP-A 115
QY 114 AVPTLPVAPGESSRAITVGANGCGSSREHAPVALGAGARATVAGGARLFFNSV 173
DB 116 SKREKVPQGEKTKYTIITIGENFGCCSSREHAPVCGAAGAAVAAGSTARLFFNSV 175
QY 174 AGEVYPLELTDVGAKKECKTGVTVDL--ANSYFINHTSGEKYKLPJGDAGVIEAG 221
DB 176 AGEVYPLD--SEVRVDCDECTGDVATVBLRBDSTILNHTGKEYKLPJGDAGVIDAG 234
QY 232 GTFAYARKGTMIASKKAA 248
DB 235 GTFAYARKGMTIPSA 251

RESULT 2

T47781
3-isopropylmalate dehydratase-like protein (small subunit) - Arabidopsis thaliana
N:Alternate names: protein F17J16.40
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 20-Apr-2000

C:Accession: T47781
R:D'Angelo, M.; Vezzi, A.; Modesto, D.; Pigazzi, M.; Valle, G.; Mewes, H.W.; Rudd, S.; I
submitted to the Protein Sequence Database, April 2000
A:Reference number: Z24476
A:Accession: T47781
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-253 <DAN>
A:Cross-references: EMBL:AL163527
A:Experimental source: cultivar Columbia; BAC clone F17J16
C:Genetics:
A:Map position: 3
A:Note: F17J16.40

Query Match	51.4%;	Score 651;	DB 2;	Length 253;
Best Local Similarity	59.7%;	Pred. No. 3.1e-46;		
Matches 135;	Conservative 29;	Mismatches 56;	Indels 6;	Gaps 4

[illegible]

RESULT 3
 A84862
 3-isopropylmalate dehydratase, small subunit [imported] - Arabidopsis thaliana
 C.Species: Arabidopsis thaliana (mouse-ear cress)
 C.Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
 C.Accession: A84862
 R.Lin. X.: Kaul, S.; Rounsley, S.D.; Shea, T.P.; Beilto, M.I.; Town, C.D.; Fujiki, C.Y.;
 W.; Koo, H.; Motilal, K.S.; Cronin, L.A.; Shen, M.; Yankem, S.E.; Unayam, J.; Tallon,
 euss, D.; Niernann, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter,
 Nature 402, 761-766, 1999
 A.Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
 A.Reference number: A84420; MUID:20083487; PMID:10617197
 A.Accession: A84862
 A.Status: preliminary
 A.Molecule type: DNA
 A.Residues: 1-256 <STO>
 A.Cross-references: GB:AE002093; NID:g3763919; PIDN:AA64259.1; GSPDB:GM00139
 C.Genetics:
 A:Gene: At2g43100
 A:Map position: 2

Query Match	49.5%;	Score 627;	DB 2;	Length 256;
Best Local Similarity	54.2%;	Pred. No. 3e-44;		
Matches 136;	Conservative 38;	Mismatches 65;	Indels 12;	Gaps 6;

0y 4 ALSGAVSTPAALAIAPRTAPTSFRIQSOTQHLRLSLTKCRAG----SIYPAAAAAGS 59

Db 11 ALPCSTYKTSSSLATFRSD---ELFNFNSTLISSISITSIRGSSPTIIPRAAASDSDS 67

0y 60 SSP-SSAVFHGCEPFWVGGNDIDTDQIIIPAEHLTIAPSPKDEYRKLGSPFAGLPSAAYPTP 118

Db 68 NMAIANTPTTHGCIYLYLKNIDTDQIIIPGAACTEPPSNQGEDEITAAHLSLPD-FIKTR 126

0y 119 FYVAGEESSRYAIIIVGANGFGCGSSREHAPVALGAAGARIVAEGRVRIEFRNSVATGEV 178

Db 127 FIEPGENRKSYSIIIGENFGCGSSREHAPVCLGAAGAKAIVAEGRVRIEFRNSVATGEV 186

0y 179 YPLELTDVGAMECKTGDDVTVYDLANS--VFIIINTSGKEYLAKIIGDAGPIVIEAGGIFAT 236

```

Db      187  FPLE -SEVAVCECKTGTVTIELSDSGGLTHTTGKNTKLKSGDAGPYIDAGGIFAY 245
QY      237  ARRTGMIAASKA 247
Db      246  ARRMGMITPSLA 256

```

RESULT 4
A72363
3-isopropylmalate dehydratase, small subunit - *Thermotoga maritima* (strain MSB8)

C:Species: *Thermobifidobacterium*
C:Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 21-Jul-2000
C:Accession: A72363
C:Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; Hickey, J.V.; Holt, J.D.; McQuinn, K.; Meyer, J.J.; Nelson, K.E.; Nelson, W.L.; Paulsen, O.; Rasmussen, S.B.; Scaife, D.; Stewart, J.; Tice, D.; White, O.; Garret, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, C.M.
Nature 399, 323-329, 1999
A:Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome
A:Reference number: A72200; MUID:99287316; PMID:10360571
A:Accession: A72363
A:Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-166 <ARN>
A:Cross-references: GB:AE001730, GB:AE000512, NID:94981062; PIDN:AD35640.1; PID:9498
A:Experimental source: strain MSB8
C:Genetics:
A:Gene: TM0555

query match

Matches	72;	Conservative	26;	Mismatches	54;	Indels	23;	Gaps
---------	-----	--------------	-----	------------	-----	--------	-----	------

```
Qy      69 GECFVVGNDNIDTDQIIPAEHLITLVPSKPEDEYRKLGSAFAAGLPSAAPTPFVAHG---EE 122
       ||| |||||:::||| | | | | | | | | | | | | | | | | | | | | | | | | |
Db      7 GKVFVEGDNVNDEIIPARYLN--TSDPQLAK-----YCMEDARPGFGGRD 51
```

```

Oy      126  SSRVAIVGGANFGCGSSREHAPVALGAAGARAIVAEGVARIFFRNSVATGEVPLELTD 185
          : ||| | ||||| : || :: : ||||| : | : :
Db      52  DIKGLIIVAGENFGCGSSREHAPVAIKAAIGISCVIAKSFARIFFRNAINIG---LPIVE 107

```

Qy 186 VGAWKECKTGDVVTVDLANSVFINTHTSGEYKCLKPKPGD-AGPVIAGGIFAIRAK 239
: | : | | | | | | | : : | | |
Db 108 LKAEDEFESCDIAEVDLNGVARNLTCKGEYRIRPPEFLMKINEAGGWLEYCLK 162

RESULT 5

C:species: Archaeoglobus fulgidus
C:date: 05-Dec-1997 #sequence_revision 05-Dec-1997
C:accession: E65328

., Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton G.G.; Gill, S.; Kirkness, E.; Glodet, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L. Nature 390, 364-370, 1997

A:Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archaeon, *Ignicoccus hospitalis*, strain 122
A:Reference number: A69250, MUID:98049343, PMID:9389475

A:Accession: G9326
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-161 <KLE>
A:Cross-references: GB:AE001061; GB:AE000782; NID:g2689384; PIDN:AAB90610.L; PID:g266

Query Match 21.58; Score 272; DB 2; Length 161;

Matches 64; Conservative 30; Mismatches 58; Indels 20; Gaps 6;

```

QY      69  GECFVVGNDITDQIIPAEHLTLVSKPDEYRKLGSEAFAGL-PSAAIPTPFVAAGEESS 12
      | : ||||| || :: : : : | | | | : |
Db      2  GRAAKFGDIDITVLIQSKLYI--NEPEE--LAKHYENLKPFA-----KEVK 47

```


GenCore version 5.1.4-p5-4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 31, 2003, 09:05:28 ; Search time 25 Seconds

(without alignments)
413.104 Million cell updates/sec

Title: US-10-027-450-47

Perfect score: 1266

Sequence: 1 MAAALSTAVSTALALPIR.....AGIPAVARKTGMASKAA 249

Scoring table: BLOSUM62

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	311.5	24.6	166	LEUD_THEMA	09w225 thermotoga
2	272	21.5	161	LE23_ARCEU	029626 archaeoglob
3	271	21.4	168	LEUD_AOUAE	067399 aquifex ae
4	267.5	21.1	168	LE24_METUA	058673 methanococ
5	266	21.0	170	LE23_METTH	026917 methanobact
6	252.5	19.9	162	LE24_METTH	027440 methanobact
7	250.5	19.8	170	LE23_METUA	038667 methanobact
8	249	19.7	165	LE24_ARCEU	028513 archaeoglob
9	202.5	16.0	163	LEUD_PYRHO	059393 pyrococcus
10	191.5	15.1	191	LEUD_LACLA	002144 lactococcus
11	189	14.9	215	LEUD_AZOVI	096196 azotobacter
12	184	14.5	744	LEUD_PHYBL	018250 phycomyces
13	183.5	14.5	758	LEUD_SCHPO	014289 schizosach
14	180	14.2	750	LEUD_RHINI	053811 rhizopus ni
15	178	14.1	199	LEUD_BACSU	094568 bacillus ni
16	176	13.9	773	LEUD_USTMA	046601 ustilago ma
17	172	13.6	200	LEUD_SYNY3	074207 synechocyst
18	171.5	13.5	208	LEUD_ALCEU	044022 alcaligenes
19	171	13.5	201	LEUD_THERH	094204 thermus the
20	165	13.0	770	LEUD_CANNA	000464 candida mal
21	164.5	13.0	755	LEUD_RHIPU	055251 rhizomucor
22	155	12.2	201	LEUD_SALTY	004787 salmonella
23	155	12.2	779	LEUD_YEAST	007264 saccharomyc
24	153.5	12.1	201	LEUD_ECOLI	030126 escherichia
25	151	11.9	200	LEUD_MAEIN	044438 haemophilus
26	144.5	11.4	693	LEUD_YEAST	044367 saccharomyc
27	139.5	11.0	197	LEUD_STRO	086535 streptomyce
28	138.5	10.9	137	LEUD_NEIRA	050181 neisseria i
29	133.5	10.5	100	LEUD_CLOPA	031960 clostridium
30	133	10.4	207	LEUD_BUCPP	094225 buchnera ap
31	131.5	10.4	887	ACOC_CABEL	023500 caenorhabdi
32	129	10.2	775	LEUD_DMENT	094412 emericella
33	128	10.1	198	LEUD_MYCLE	033124 mycobacteri

34	128	10.1	198	LEUD_MYCTU	053236 mycobacteri
35	127	10.0	207	LEUD_BUCAI	056935 buchnera ap
36	126	10.0	208	LEUD_BUCDN	085073 buchnera ap
37	125	9.9	961	ACON_MYCAV	008451 mycobacteri
38	122	9.6	207	LEUD_BUCAP	085066 buchnera ap
39	122	9.6	207	LEUD_BUCRP	048574 buchnera ap
40	120.5	9.5	205	LEUD_BUCTS	031294 buchnera ap
41	118	9.3	891	ACON_LEGPN	037032 legionella
42	116.5	9.2	780	ACON_BOVIN	020004 bos taurus
43	116	9.2	878	ACON_RICCN	092950 rickettsia
44	115.5	9.1	780	ACON_HUMAN	099798 homo sapien
45	114.5	9.0	781	ACON_PIC	016276 sus scrofa

ALIGNMENTS

RESULT 1	ID	LEUD_THEMA	STANDARD:	PRT:	166 AA.
AC	09w225	30-MAY-2000 (Rel. 39, Created)			
DT	30-MAY-2000 (Rel. 39, Last sequence update)				
DT	15-JUN-2002 (Rel. 41, Last annotation update)				
DE	3-isopropylmalate dehydratase small subunit (EC 4.2.1.33)				
DE	(Isopropylmalate isomerase) (Alpha-IPM isomerase) (IPMI).				
GN	LEUD OR TM0555.				
OS	Thermotoga maritima.				
OC	Bacteria; Thermotogae; Thermotogae (class); Thermotogales;				
OC	Thermotogaceae; Thermotoga.				
OX	NCBI_TaxID=2336;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=MSB8 / DSM 3109;				
RX	MEDLINE=99287316; PubMed=10360571;				
RA	Nelson K.E., Clayton R.A., Gill S.R., Gwin M.L., Dodson R.J.,				
RA	Halt D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,				
RA	McDonald L., Utterback T.R., Malek J.A., Linher K.D., Garrett M.M.,				
RA	Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.,				
RA	Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,				
RA	Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.;				
RL	"Evidence for lateral gene transfer between Archaea and Bacteria from genome sequence of Thermotoga maritima.";				
CC	Nature 399:323-329(1999).				
CC	-1- CATALYTIC ACTIVITY: 3-ISOPROPYLMALATE = 2-ISOPROPYLMALATE + H(2)O (ALSO CATALYZES 2-ISOPROPYLMALATE + H(2)O = 3-HYDROXY-4-METHYL-3-CARBOXY-PENTANONE).				
CC	-1- PATHWAY: leucine biosynthesis; second step.				
CC	-1- SUBUNIT: CONSISTS OF TWO DIFFERENT SUBUNITS: LEUC AND LEUD.				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).				
CC	or send an email to license@isb-sib.ch).				
DR	EMBL: AE001750; AAD35640.1; -.				
DR	TIGR: TM0555; -.				
DR	InterPro: IPR000573; Aconitase_C.				
DR	Pfam: PF00694; Aconitase_C; 1.				
KW	leucine biosynthesis; lyase; Complete proteome.				
SO	SEQUENCE 166 AA; 18451 MW; 8536DCB6B0D22F.CRC64;				
QY	69 GECFVAGNITDQITPAHEHTLVPSKRPDEYRKGSFAAGLPSAAYPTFPVAGG---EE 125				
DB	7 GKVFVGGNVTMTDEIPARYLN--TSDFOELAK-----YCMEDARGGGRRD 51				
	Query Match 24.6%; Score 311.5; DB 1; Length 166;				
	Best Local Similarity 41.1%; Pred. No. 1e-18;				
	Matches 72; Conservative 26; Mismatches 54; Indels 23; Gaps 5;				

[illegible]

RESULT 2			
LE23_ARCFU			
ID	LE23_ARCFU	STANDARD:	PRT; 161 AA.
AC	029626;		
DT	15-JUL-1998 (Rel. 36, Created)		
DT	15-JUL-1998 (Rel. 36, Last sequence update)		
DT	15-JUN-2002 (Rel. 41, Last annotation update)		
DE	Putative 3-isopropylmalate dehydratase small subunit (EC 4.2.1.33)		
DE	(isopropylmalate isomerase) (Alpha-1PM isomerase) (IPMI).		
GN	AF0659.		
OS	Archaeoglobus fulgidus.		
OC	Archaea; Euryarchaeota; Archaeoglobi; Archaeoglobales;		
OC	Archaeoglobaceae; Archaeoglobus.		
OX	NCBI_TaxId=2234;		

RP SEQUENCE FROM N.A.
RC STRAIN=VC-16 / DSM 4304 / ATCC 49558;
RX MEDLINE=98049343; Pubmed=9339475;
RA Klein H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,
Raetchum K.A., Dodson R.U., Gwinn M., Hickey E.K., Peterson J.D.,
RA Richardson D.L., Kerlavage A.R., Graham D.E., Kyrtides N.C.,
RA Fleischmann D.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,
RA Kirkness E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,
RA Peterson S., Reich C.I., McNeil L.K., Badger J.H., Glodek A., Zhou L.,
RA Overbeek R., Gocayne J.D., Weidman J.F., McDonald L., Utterback T.,
RA Cotton M.D., Spriggs T., Artzt P., Kaine B.P., Sykes S.M.,
RA Sedow P.W., D'Andrea K.P., Bowman C., Fujii C., Gialand S.A.,
RA Mason T.M., Olsen G.U., Fraser C.M., Smith H.O., Woese C.R.,
RA Venter J.C.,

RT The complete genome sequence of the hyperthermophilic, sulphate
 RT reducing archaeon *Archaeoglobus fulgidus*.",
 RL Nature 390:364-370(1997).
 CC -1- CATALYTIC ACTIVITY: 3-ISOPROPYLMALATE = 2-ISOPROPYLMALATE +
 CC H(2)O (ALSO CATALYZES 2-ISOPROPYLMALATE + H(2)O = 3-HYDROXY-
 CC 4-METHYL-3-CARBOXPENTANONE).
 CC -1- PATHWAY: Leucine biosynthesis; second step.

- SUBUNIT: CONSISTS OF TWO DIFFERENT SUBUNITS POTENTIAL).

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce> or send an email to license@isb-sib.ch).

CC
DR EMBL, AE001061, AAB90610.1, -,
DR HSSP, P20004, 1AMJ,
DR TIGR, AF0629, -,
DR InterPro, IPR000573, Aconitase_C,
DR Pfam, P00694, Aconitase_C, 1,
DR Leucine biosynthesis, Lyase, Complete proteome,
KW SEQUENCE 161 AA; 18137 MW; 6DEB0A2AE8B1CA7
CRC64;

Query Match	21.5%;	Score 272;	DB 1;	Length 161;
Best Local Similarity	37.2%;	Pred. No. 1.6e-15;		
Matches 64;	Conservative 30;	Mismatches 58;	Indels 20;	Gaps 6

```

QY      69  GGCFTVGNINPTDQIIPAEHTLVTPSPKPDYKRLKSGFAAGL-PSAAIPTFPVARGCESS 127
          | : ||| ||| || : : : ||| : : : ||| : : : ||| : : : ||| : : : |||
Db      2  GRAMFEGDDITDVTYIGKYLVI--NPEP-----LAKHVEENLRPEFA-----KEVK 47
          | : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY     128  RAILIVGANGCGGSSRHHAPVALGAAGAAIYAEGYARIFFRNSVATGEVYPLETIDVG 187
          : : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

Dd 48 KGDFVAVGEGNCGSSREHPALAKANGIAEIVAIKASTARLFFRNNAITG---LRYLECK 103
Qy 188 AMKEKTGDVTVDLANSVFNFHTSGAEYKIKPIGD-AGVIEAGGIFANAR 238
 : : : : : | : : : : : | : : : : :
Dd 104 ETDRIEDDELEVYEKGVIYNKKGGEEYIPNPDPDLKLLELKGGLVEPAK 155

LEUD_3	LEUD_AQUAE	STANDARD:	PRT:	168 AA.
AC	067399,			
DT	30-MAY-2000 (Rel. 39, Created)			
DT	30-MAY-2000 (Rel. 39, Last sequence update)			
DT	15-JUN-2002 (Rel. 41, Last annotation update)			
DE	3-isopropylmalate dehydratase small subunit (EC 4.2.1.33)			
DE	(Isopropylmalate isomerase) (Alpha-IPM isomerase) (IPMI).			
GN	LEUD OR AC.1398.			
OS	Aquifex zeolites			
OC	Bacteria: Aquificae; Aquificae (class); Aquificales; Aquificaceae;			
CC	Aquifex.			
XX	NCBI_TaxID=63363;			

RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=VF5;
 RX MEDLINE=98196666; PubMed=9537320;
 RA Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,
 RA Graham D.E., Overbeek R., Sneed M.A., Keller M., Aufay M., Huber R.,
 RA Feldman R.A., Short J.M., Olson G.J., Swanson R.V.;
 RT "The complete genome of the hyperthermophilic bacterium *Aquifex*
 aeolicus.";
 RL Nature 392:353-358(1998).
 CC -1- CATALYTIC ACTIVITY: 3-ISOPROPYLMALATE = 2-ISOPROPYLMALATE +
 CC H₂O (ALSO CATALYZES 2-ISOPROPYLMALATE + H₂O = 3-HYDROXY-
 CC 4-METHYL-3-CARBOXY-PENTANONE).

CC -1- PAYHWA: Leucine biosynthesis; second step.
CC -1- SUBUNIT: CONSISTS OF TWO DIFFERENT SUBUNITS: LEUC AND LEUD.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement>
CC or send an email to license@sib-sib.ch).
CC
DR EMBL: AEO00738; AAC07359.1; AAT_INT1.
DR InterPro: IPR000573; Aconitase_C.
DR Pfam: PF00694; Aconitase_C.1.
KW Leucine biosynthesis; Lyase; Complete proteome.
KW SOURCE 168 AA; 18786 MW; 25A3AFEF7FCA09F3 CRC64;
KW

Query Match	21.48;	Score 271;	DB 1;	Length 168;
Best Local Similarity	37.58;	Pred. No. 2.1e-15;		
Matches 72;	Conservative 27;	Mismatches 53;	Indels 40;	Gaps 8;

0y 67 FHEGCEFWGNDIDTQIIPAEHLLTVSPKDEYRKLGSFAAGLPSAAYPIPFVARGES 120
| | : |||:||||||| :| | | |
Db 5 FRGFWKFGNDVDQIIPARYLN-TSDPYELAK-----HYMEDSEH 45

127 SRYA-----IIVGANFGCGSSREHAPVALGAACARAIVAEGYARIFFRNSVATGEVY 179

QY 180 PLEETDV-GAMKECTGDVVTVDLANSVEFINHTSGKEYKL---KPIGDAGPVIACGIF 234A
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
db 103 -LDIVAEAPVDIEFGDETVDI.EKGCITKINI.BTCKEYVAATKEKETOD---TI.KACIM 155

QY	235	AYARKTGMIASK	246
	::	::	
Db	159	AYAKE--KLASK	168

RESULT 4

```

LE24_METTA STANDARD; PRT; 168 AA.
ID LE24_METTA
AC 058673:
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DE 15-JUN-2002 (Rel. 41, Last annotation update)
DE Putative 3-isopropylmalate dehydratase small subunit (EC 4.2.1.33)
DE (isopropylmalate isomerase) (Alpha-IPM isomerase) (IPMI).
GN MJ1277.
OS Methanococcus jannaschii.
OC Archaea: Euryarchaeota: Methanococci: Methanococcales;
OC Methanocaldococcaceae; Methanocaldococcus.
OX NCBI_TaxID=2190;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-JAL-1 / DSM 2661 / ATCC 43067;
RX MEDLINE=96337999; PubMed=6888087;
RA Sult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Sult G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
RA Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA Scott J.L., Geoghegan N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
RA Uitterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA Clontz H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RA "Complete genome sequence of the methanogenic archaeon, Methanococcus
RT jannaschii."
RL Science 273:1058-1073(1996).
CC -1- CATALYTIC ACTIVITY: 3-ISOPROPYLMALATE = 2-ISOPROPYLMALATE +
CC H(2)O (ALSO CATALYSES 2-ISOPROPYLMALATE + H(2)O = 3-HYDROXY-
CC 4-METHYL-3-CARBOXPENTANONE).
CC -1- PATHWAY: leucine biosynthesis; second step.
CC -1- SUBUNIT: CONSISTS OF TWO DIFFERENT SUBUNITS (POTENTIAL).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U67568; AAB9283.1; -.
CC TIGR: MJ1277; -.
DR InterPro: IPR000573; Aconitase_C.
DR Pfam: PF00694; Aconitase_C.1.
KM Leucine biosynthesis; Lyase; Complete proteome.
SQ SEQUENCE 168 AA; 18377 MW; A53C2A883B6C07 CRC64;

Query Match 21.1%; Score 267.5; DB 1; Length 168;
Best Local Similarity 38.3%; Pred. No. 4e-15;
Matches 69; Conservative 29; Mismatches 57; Indels 25; Gaps 9;

QY 65 AVFGCEFYVGNIDTDQIIPAEHLLTVPSKPDVEYRKLSGFAFAGLPSAAYPTFFVAPGE 124
DB 3 STIKGVWMEFGNVDTDILPARY--LYYTKPEE--LAQFVYMG-ADPDPRK-VARGD 55
QY 125 ESSRAIIVYGANFGCGSSREHAPVALGAGARAIYAEGRARIFRNSVATGEVYPLELT 184
DB 56 -----IIVAGRNFGCGSSREHAPGLGAGISCVIAESFARIFYRNAING--LPL-IE 106
QY 185 DYVAMKECKTGDPVYVDLANSVFINTSGKEYKLPIDGAP-----VIEAGIGFAVARK 239
DB 107 CAGISKRVNMGDELEVNLETGEIKNLTLTGEVYK-----GQKPEFMETLEAGGIMPLYLKK 162

```

```

DE Putative 3-isopropylmalate dehydratase small subunit (EC 4.2.1.33)
DE (isopropylmalate isomerase) (Alpha-IPM isomerase) (IPMI).
GN M7829.
OS Methanobacterium thermoautotrophicum.
OC Archaea: Euryarchaeota: Methanobacteria: Methanobacteriales;
OC Methanobacteriaceae; Methanothermobacter.
OX NCBI_TaxID=187420;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-Delta H.
RX MEDLINE=96037514; PubMed=9371463;
RA Smith D.R., Doucette-Stamm L.A., Delonghery C., Lee H.-M., Dubois J.,
RA Aldredge T., Bashirzadeh R., Blakey D., Cook R., Gilbert K.,
RA Harrison D., Hoang L., Keagle P., Lumm W., Pothier B., Qiu D.,
RA Spadafora R., Vlcare R., Wang Y., Wierzbowski J., Gibson R.,
RA Jiweni N., Caruso A., Bush D., Safer H., Patwell D., Prabhakar S.,
RA McPougall S., Shimer G., Goyal A., Pietrowski S., Church G.M.,
RA Daniels C.J., Mao J.-I., Rice P., Noelling J., Reeve J.N.;
RT "Complete genome sequence of Methanobacterium thermoautotrophicum
RT deltaH: functional analysis and comparative genomics."
CC -1- CATALYTIC ACTIVITY: 3-ISOPROPYLMALATE = 2-ISOPROPYLMALATE +
CC H(2)O (ALSO CATALYSES 2-ISOPROPYLMALATE + H(2)O = 3-HYDROXY-
CC 4-METHYL-3-CARBOXPENTANONE).
CC -1- PATHWAY: leucine biosynthesis; second step.
CC -1- SUBUNIT: CONSISTS OF TWO DIFFERENT SUBUNITS (POTENTIAL).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AE00860; AAB85327.1; -.
DR InterPro: IPR000573; Aconitase_C.
DR Pfam: PF00694; Aconitase_C.1.
KM Leucine biosynthesis; Lyase; Complete proteome.
SQ SEQUENCE 170 AA; 18651 MW; BC07146AF7585215 CRC64;

Query Match 21.0%; Score 266; DB 1; Length 170;
Best Local Similarity 39.5%; Pred. No. 5.4e-15;
Matches 70; Conservative 24; Mismatches 61; Indels 22; Gaps 7;

QY 66 VFHGCIFYVGNIDTDQIIPAEHLLTVPSKPDVEYRKLSGFAFAGLPSAAYPTFFVAPGE 124
DB 4 IIRGVWNRGVDVDDMIIPGRYL-----RFFSDELASHVMEGARREFRSQ--VRKGD 55
QY 125 ESSRAIIVYGANFGCGSSREHAPVALGAGARAIYAEGRARIFRNSVATGEVYPLELT 184
DB 56 -----IIVAGRNFGCGSSREDAVALKHAQVVAIIAESFARIFYRNAING--LPIYMA 107
QY 185 DYVAMKECKTGDPVYVDLANSVFINTSGKEYKLPIDG-AGPYIEAGIGFAVARKT 240
DB 108 KYDA-----DDGDEVIDLRSGQIRNLTAGSEYRKPFNDYMLSLDEGGGLVNHLYKT 160

RESULT 6
LE24_METTH STANDARD; PRT; 162 AA.
ID LE24_METTH
AC 027440:
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DE 15-JUN-2002 (Rel. 41, Last annotation update)
DE Putative 3-isopropylmalate dehydratase small subunit (EC 4.2.1.33)
DE (isopropylmalate isomerase) (Alpha-IPM isomerase) (IPMI).
GN MTH1387.
OS Methanobacterium thermoautotrophicum.
OC Archaea: Euryarchaeota: Methanobacteria: Methanobacteriales;
OC Methanobacteriaceae; Methanothermobacter.
OX NCBI_TaxID=187420;
RN [1]

```



```

CC 4-METHYL-3-CARBOXPENTANONE).
CC -1- PATHWAY: Leucine biosynthesis; second step.
CC -1- SUBUNIT: CONSISTS OF TWO DIFFERENT SUBUNITS: LEUC AND LEUD.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U92974; AAB81916.1; -.
CC DR EMBL: AE006354; AAK05319.1; -.
CC DR PIR: S35135; S35135.
CC DR InterPro: IPR000573; Aconitase_C.
CC DR InterPro: IPR004431; Leud.
CC DR Pfam: PF00694; Aconitase_C.1.
CC DR TIGRfams: TIGR00171; Leud; 1.
CC DR Leucine biosynthesis; Lyase; Complete proteome.
CC KW Leucine biosynthesis; Lyase; Complete proteome.
CC SEQUENCE 191 AA; 21849 MW; C691BC6AEEF4E964 CRC64;

Query Match 15.1%; Score 191.5; DB 1; Length 191;
Best Local Similarity 29.0%; Pred. No. 7.3e-09;
Matches 58; Conservative 35; Mismatches 70; Indels 37; Gaps 7;

QY 58 GSSPSAFAVFGEVVDNIDTDQIIPAEHLTVPSKRPDEYKLSFAFAGLPSAAYPT 117
DB 9 GTSPV-----VAMNDNIDTDQIIPKQKLAIDK--GFGKNLPEYEMRYLKDYDENP 56
QY 118 PFVAPGEESRYAIIIGVANGFCSSREHAPVALGAAGARAIVAGYARIFFRNSVATGE 177
DB 57 DEILNAPKRYKKAALLISGDNFGSSREHAMALSDYGFRAIIIVGSDIFYNALANG- 115
QY 178 VYPL-----ELTDVGAKKECKTGCVYTVDLANSVFINTSKREKLPKPI----- 221
DB 116 LILPIKPREVINQLTKSSQEE-----ITIDLPHQILITSLGDFHEIDIPMKDLING 169
QY 222 -GDAGFYIE-AGGIFAYARK 239
DB 170 LDDIGITLQVEBAISAYEQK 189

RESULT 11
LEUD_AZOV1 STANDARD; PRT; 215 AA.
AC P96196;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE 3-isopropylmalate dehydratase small subunit (EC 4.2.1.33)
DE (isopropylmalate isomerase) (Alpha-IPM isomerase) (IPMI).
GN LEUD.
OS Azotobacter vinelandii.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Azotobacter.
OX NCBI_TaxID=354;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=UM;
RX MEDLINE=97261869; PubMed=9108283;
RA Manna A.C., Das H.K.;
RA "Characterization and mutagenesis of the leucine biosynthetic genes
RT of Azotobacter vinelandii: an analysis of the rarity of amino acid
RT auxotrophs.";
RL Mol. Genet. 254:207-217(1997).
CC -1- CATALYTIC ACTIVITY: 3-ISOPROPYLMALATE = 2-ISOPROPYLMALATE +
CC H(2)O (ALSO CATALYZES 2-ISOPROPYLMALATE + H(2)O = 3-HYDROXY-
CC 4-METHYL-3-CARBOXPENTANONE).
CC -1- PATHWAY: Leucine biosynthesis; second step.
CC -1- SUBUNIT: CONSISTS OF TWO DIFFERENT SUBUNITS: LEUC AND LEUD.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

```

```

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: Y11280; CA72150.1; -.
CC DR EMBL: Y11280; CA72150.1; -.
CC DR InterPro: IPR000573; Aconitase_C.
CC DR InterPro: IPR004431; Leud.
CC DR Pfam: PF00694; Aconitase_C.1.
CC DR TIGRfams: TIGR00171; Leud; 1.
CC DR Leucine biosynthesis; Lyase.
CC KW Leucine biosynthesis; Lyase.
CC SEQUENCE 215 AA; 24133 MW; 8E7CC713228005B8 CRC64;

Query Match 14.9%; Score 189; DB 1; Length 215;
Best Local Similarity 35.2%; Pred. No. 1.3e-08;
Matches 56; Conservative 26; Mismatches 59; Indels 18; Gaps 8;

QY 77 NIDTDQIIPAEHLTV-----PSKRPDEYKLSFAFAGLPSAAYPT-TPVAPGEESR 128
DB 18 NVDTQDILIPKPLKSLKRTGFGPNLDEWRYL-DVGPGQDCSARPLMTGTYVNLPRYG 76
QY 129 YAIIVGANGFCSSREHAPVALGAAGARAIVAGYARIFFRNSVATGEVYPLDLYGA 188
DB 77 ASYLAKRENFQCGSSREHAPVALDEGFRFTVAPSFADIFFNNSPKNG-LPLILPEAEV 135
QY 189 ---WKECKT--GDVYTVDLANSVFINTSKG--EYKLP 220
DB 136 DELFRCCESTEGYRLTVDLAAQT-VTRPDGKALSEIDP 173

RESULT 12
LEUD_PHYBL STANDARD; PRT; 744 AA.
AC P18250;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE 3-isopropylmalate dehydratase (EC 4.2.1.33) (isopropylmalate
DE isomerase) (Alpha-IPM isomerase) (IPMI).
GN LEU1.
OS Phycomyces blakesleeanus.
OC Eukaryota; Fungi; Zygomycota; Zygomycetes; Mucorales; Mucoraceae;
OC Phycomyces.
OX NCBI_TaxID=4837;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NRRL 1555;
RX MEDLINE=90356419; PubMed=238845;
RA Iturriaga E.A., Diaz-Minguez J.M., Benito E.P., Alvarez M.I.,
RA Esalva A.P.;
RA "Nucleotide sequence of the phycomyces blakesleeanus leu1 gene.";
RL Nucleic Acids Res. 18:4612-4612(1990).
GN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=NRRL 1555;
RX MEDLINE=92224296; PubMed=1563047;
RA Iturriaga E.A., Diaz-Minguez J.M., Benito E.P., Alvarez M.I.,
RA Esalva A.P.;
RA "Heterologous transformation of Mucor circinelloides with the
RT phycomyces blakesleeanus leu1 gene.";
RL Curr. Genet. 21:215-223(1992).
RN [3]
RP IDENTIFICATION OF PROBABLE FRAMESHIFT.
RA Gibson T.J.;
RL Unpublished observations (MAR-1996).
CC -1- FUNCTION: Catalyzes the isomerization between 2-isopropylmalate
CC and 3-isopropylmalate, via the formation of 2-isopropylmaleate.
CC -1- CATALYTIC ACTIVITY: 3-isopropylmalate = 2-isopropylmaleate +
CC H(2)O.
CC -1- CATALYTIC ACTIVITY: 2-isopropylmaleate + H(2)O = 2-
CC isopropylmalate.

```



```

CC -1- PATHWAY: Leucine biosynthesis; second step.
CC -1- SUBUNIT: Monomer.
CC -1- SIMILARITY: BELONGS TO THE ACONITASE/IPM ISOMERASE FAMILY.
CC -1- CAUTION: Ref.1 and Ref.2 sequences differ from that shown due to a
CC frameshift in position 676
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL: X53090; CAA37257.1; ALT_FRAME.
CC PIR: S10998; S10998.
CC PIR: S26864; S26864.
CC InterPro: IPR000573; Aconitase_C.
CC InterPro: IPR001030; Aconitase_N.
CC InterPro: IPR004430; Aconitase_N.
CC InterPro: IPR004431; Leuc.
CC Pfam: PF00330; aconitase; 1.
CC Pfam: PF00694; Aconitase_C.1.
CC PRINTS: PR00415; Aconitase.
CC ProDom: PD000511; Aconitase_N.1.
CC TIGRFS: TIGR00170; leuc; 1.
CC TIGRFS: TIGR00171; leud; 1.
CC PROSITE: PS00450; ACONITASE_1; 1.
CC PROSITE: PS01244; ACONITASE_2; 1.
CC Leucine biosynthesis: Lyase: Iron-sulfur; 4Fe-4S.
CC METAL: 341 341 IRON-SULFUR (4FE-4S) (BY SIMILARITY).
CC METAL: 401 401 IRON-SULFUR (4FE-4S) (BY SIMILARITY).
CC METAL: 404 404 IRON-SULFUR (4FE-4S) (BY SIMILARITY).
CC SEQUENCE 744 AA; 80833 MW; 9E5A93089103745C CRC64;

Query Match 14.5%; Score 184; DB 1; Length 744;
Best Local Similarity 38.0%; Pred. No. 1.4e-07;
Matches 57; Conservative 15; Mismatches 62; Indels 16; Gaps 8;

OY 42 CRRASIVPAAAMAAAGSSPSAVFHGECFVG-DNIDTQIIPAEHLTVPSKPEYR 100
DB 499 CRRPAGS--PVNKGAPVANAAMRAFTLLKGVAPLAISNVDIMILPKQFL-----KTKRT 551
OY 101 KLGSAFAGL--PSAAVTP-ETV--ADGESSRYAIIIGVANGCGSSREHAPVALGAA 154
DB 552 GIGSALFVGLRYDPATGAKEKPEFLNQPAYRSSK-ILVCTGPNMGCGSSREHAPVANDF 610
OY 155 GARAIYAEAGYARIFFRNSVATGEVYPLELT 184
DB 611 GIRCIITATSFADIFFNRCFKNG-MLPITLS 639

RESULT 13
LEU2.SCHPO STANDARD: PRT; 758 AA.
AC 014289:
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DE 15-JUN-2002 (Rel. 41, Last annotation update)
DE 3-isopropylmalate dehydratase (EC 4.2.1.33) (isopropylmalate
DE isomerase) (Alpha-IPM isomerase) (IPMI).
GN SPAC9E9.03.
OS Schizosaccharomyces pombe (fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RX MEDLINE=21848401; Pubmed=11859360;
RA Wood V., Gilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA Sgouras J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,

```

```

RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Felwell T., Fraser A.,
RA Gentsch S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jajels K.,
RA James K., Jones L., Jones M., Leather S., McConall S., McLean J.,
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., O'Neill C.,
RA Olliver K., O'Neill S., Pearson D., Quail M.A., Rabbittsch E.,
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA Woodward J., Voicakeart G., Aert R., Robben J., Grymoprez B.,
RA Willems I., Vansteels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,
RA Borzym K., Langer I., Beck A., Lehnach H., Reinhardt R., Pohl T.M.,
RA Eger P., Zimmermann W., Medler E., Dreano S., Gloux S., Lelaure V., Mottier S.,
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Moore K., Hurst S.M.,
RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA Lucas M., Rochet M., Galliard C., Tallada V.A., Garzon A., Thode G.,
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA Cerutti L., Lowe T., McCombie W.R., Paulsen O., Potashkin J.,
RA Shpakovskii G.V., Usero D., Barrell B.G., Nurse P.,
RA The genome sequence of Schizosaccharomyces pombe.
RA Nature 415:871-880 (2002).
CC -1- FUNCTION: Catalyzes the isomerization between 2-isopropylmalate
CC and 3-isopropylmalate, via the formation of 2-isopropylmalate.
CC -1- CATALYTIC ACTIVITY: 3-isopropylmalate = 2-isopropylmalate +
CC H(2)O.
CC -1- CATALYTIC ACTIVITY: 2-isopropylmalate + H(2)O = 2-
CC isopropylmalate.
CC -1- PATHWAY: Leucine biosynthesis; second step.
CC -1- SIMILARITY: BELONGS TO THE ACONITASE/IPM ISOMERASE FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL: Z99262; CAB16402.1; -.
CC InterPro: IPR000573; Aconitase_C.
CC InterPro: IPR001030; Aconitase_N.
CC InterPro: IPR004430; Leuc.
CC InterPro: IPR004431; Leud.
CC Pfam: PF00330; aconitase; 1.
CC Pfam: PF00694; Aconitase_C.1.
CC PRINTS: PR00415; Aconitase.
CC ProDom: PD000511; Aconitase_N.1.
CC TIGRFS: TIGR00170; leuc; 1.
CC TIGRFS: TIGR00171; leud; 1.
CC PROSITE: PS00450; ACONITASE_1; 1.
CC PROSITE: PS01244; ACONITASE_2; 1.
CC Leucine biosynthesis: Lyase: Iron-sulfur; 4Fe-4S.
CC METAL: 359 359 IRON-SULFUR (4FE-4S) (BY SIMILARITY).
CC METAL: 420 420 IRON-SULFUR (4FE-4S) (BY SIMILARITY).
CC METAL: 423 423 IRON-SULFUR (4FE-4S) (BY SIMILARITY).
CC SEQUENCE 758 AA; 82782 MW; CE78C36828380E47 CRC64;

Query Match 14.5%; Score 183.5; DB 1; Length 758;
Best Local Similarity 32.1%; Pred. No. 1.6e-07;
Matches 60; Conservative 19; Mismatches 77; Indels 31; Gaps 6;

OY 53 AAAAAAGSSPSAVFHGECFVG-----DNIDTQIIPAEHLTVPSKPEYR 105
DB 523 ATNVAGSVSSGSGIPKFTVYEGIAAPLPMANVDTKIKPKQFL-----TKIKTGIGQF 577
OY 106 AFAGLPASAYPTPEVAPGESGRY-----AIIYGANFCGSSREHAPVALGAA 155
DB 578 AF-----YETRYDADGKEIPDFVLANEPRKATVLAHNEFCGSSREHAPVALNDG 630
OY 156 ARAIYAEAGYARIFFRNSVATGEV-YPLELTDVGAAMECKETGCV-VTVLANSVYFINHTSG 213

```

DB 631 IRVIAESFADIFFNCKRMGLPIPIPIEDVNDMAKRAENQVAFSDVNVQTTTGDQ 690
OY 214 KEYLKP 220
DB 691 VKFVEP 697

RESULT 14
LEU2_RHINI STANDARD; PRT; 750 AA.
AC P55811;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE 3-Isopropylmalate dehydratase (EC 4.2.1.33) (isopropylmalate
isomerase) (Alpha-IPM isomerase) (IPMI).
GN LEU1.
OS Rhizopus niveus.
OC Eukaryota; Fungi; Zygomycota; Zygomycetes; Mucorales; Mucoraceae;
OC Rhizopus.
OX NCBI_TaxID=4844;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Yamazaki / IFO 4810;
RA MEDLINE=97056764; PubMed=8901103;
RT Takaya N., Yanai K., Horuchi H., Ohta A., Takagi M.;
RT "Cloning and characterization of the Rhizopus niveus leu1 gene and
RT its use for homologous transformation.";
RL Biosci. Biotechnol. Biochem. 60:448-452(1996).
CC -1- FUNCTION: Catalyzes the isomerization between 2-isopropylmalate
CC and 3-isopropylmalate, via the formation of 2-isopropylmalate.
CC -1- CATALYTIC ACTIVITY: 3-isopropylmalate = 2-isopropylmalate +
CC H(2)O.
CC -1- CATALYTIC ACTIVITY: 2-isopropylmalate + H(2)O = 2-
CC isopropylmalate.
CC -1- PATHWAY: Leucine biosynthesis; second step.
CC -1- SUBUNIT: Monomer (by similarity).
CC -1- SIMILARITY: BELONGS TO THE ACONITASE/IPM ISOMERASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D63833; BAA09893.1; -
DR InterPro; IPR000573; Aconitase_C.
DR InterPro; IPR001030; Aconitase_N.
DR InterPro; IPR004430; Leuc.
DR InterPro; IPR004431; Leud.
DR Pfam; PF00330; aconitase; 1.
DR Pfam; PF00694; aconitase; 1.
DR PRINTS; PR00415; ACONITASE.
DR PRODOM; PD000511; Aconitase_N.1.
DR TIGRFAMs; TIGR00170; leuc; 1.
DR TIGRFAMs; TIGR00171; leud; 1.
DR PROSITE; PS00450; ACONITASE.1; 1.
DR PROSITE; PS01244; ACONITASE.2; 1.
KM Leucine biosynthesis; Lyase; iron-sulfur; 4Fe-4S.
FT MERRL 353 353
FT MERRL 413 413
FT MERRL 416 416
SQ SEQUENCE 750 AA; 61308 MW; 7D7BE28304E26A83 CRC64;
Query Match 14.28 Score 180; DB 1; Length 750;
Best Local Similarity 34.58; Pred. No. 3e-07;
Matches 59; Conservative 20; Mismatches 72; Indels 20; Gaps 8;
OY 50 PAAAAAAGSSSPSSAVHHEGEFVVG-DNIPDDIIPPEHLTLVPSKDEVRKIGSFAPA 108
||:|||||:|:|:|||||||:|:|:|||||

DB 515 PASDSSSG-GMPAFITLKGYAAPIDISNIDDMIIKPOFL-----KTIKRTGLSALFY 568
OY 109 GL---PSAAYPPPEYVAGESSRYA--IIVGANFEGCGSSREHAPVALGACARALVAG 163
DB 569 SLRFDPQGTGAENPAFLKERTFEROARIIVCTGPNRCGSSREHAPWAFNDFGRIILAPS 628
OY 164 YARIEFRNSVANGGEVYPL-----ELTDVGAMKECKTGADVTVDLANSYFVN 209
DB 629 FADIFFNCKRMG-MLPVLPQALFAINA--EAQKGVEVDVQOIVRN 676

RESULT 15
LEUD_BACSU STANDARD; PRT; 199 AA.
AC P94568;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE 3-Isopropylmalate dehydratase small subunit (EC 4.2.1.33)
DE (isopropylmalate isomerase) (Alpha-IPM isomerase) (IPMI).
GN LEUD.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RA MEDLINE=97124191; PubMed=8969504;
RT Wipat A., Carter N., Brignell C.S., Guy J.B., Piper K.,
RA Sanders J., Emerson P.T., Harwood C.R.;
RT "The dnaB-*chroma* (236 degrees-240 degrees) region of the Bacillus
RT subtilis chromosome containing genes responsible for stress
RT responses, the utilization of plant cell walls and primary
RT metabolism.";
RT Microbiology 142:3067-3078(1996).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RA MEDLINE=98044033; PubMed=9384377;
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
RA Borriss R., Bourcier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA Brouillet S., Brusch C.V., Caldwell B., Capuano V., Carter N.M.,
RA Choi S.K., Codani J.J., Conneron I.F., Cummings N.J., Daniel R.A.,
RA Deniot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,
RA Entlin K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA Fritz C., Fujita M., Fujita Y., Fuma S., Gallizzi A., Galleron N.,
RA Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
RA Giuseppe G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,
RA Hillebert H., Holsappel S., Hosono S., Hult M.F., Itaya M., Jones L.,
RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
RA Kobayashi Y., Koelter P., Koningsstein G., Krogh S., Kumano M.,
RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,
RA Medina N., Mellado R.P., Mizuno M., Mestl D., Nakai S., Noack M.,
RA Noone D., O'Reilly M., Ogawa K., Ogihara A., Oudega B., Park S.H.,
RA Pario V., Pohl T.M., Portetlelle D., Portetlelle S., Prescott A.M.,
RA Presecan E., Puig J.M., Purnelle D., Rapoport G., Rey M., Reynolds S.,
RA Rieger M., Rivolta C., Roche E., Roche B., Rose M., Sadaie Y.,
RA Sato T., Scanlan E., Schleich S., Schoefer R., Scoffone F.,
RA Sekiguchi J., Sekowska A., Seir S.J., Serro P., Shin B.S., Soldo B.,
RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA Takenchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassaretli A.,
RA Viari A., Wambutt R., Wedler E., Wedler H., Wellzenegger T.,
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasunoto K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zimstein E., Yoshikawa H., Danchin A.;
RT "The complete genome sequence of the Gram-positive bacterium Bacillus
RT subtilis.";
RL Nature 390:249-256(1997).
CC -1- CATALYTIC ACTIVITY: 3-ISOPROPYLMALATE = 2-ISOPROPYLMALATE +
CC H(2)O (ALSO CATALYZES 2-ISOPROPYLMALATE + H(2)O = 3-HYDROXY-
CC 4-METHYL-3-CARBOXPENTANONE).


```
CC -!- PATHWAY: leucine biosynthesis; second step.
CC -!- SUBUNIT: CONSISTS OF TWO DIFFERENT SUBUNITS: LEUC AND LEUD.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: 275208; CAA9534.1; -.
DR EMBL: 299118; CAB14785.1; -.
DR Subtilist; BG11950; leud.
DR InterPro; IPR000573; Aconitase_C.
DR Pfam; PF00694; Aconitase_C; 1.
DR TIGRFAMs; TIGR00171; leud; 1.
KW leucine biosynthesis; lyase; Complete proteome.
SQ SEQUENCE 199 AA: 23022 MW: 99ADPEL2A0D5D9D3 CRC64;

Query Match 14.1%; Score 178; DB 1; Length 199;
Best Local Similarity 34.2%; Pred. NO. 9.7e-08;
Matches 53; Conservative 22; Mismatches 48; Indels 32; Gaps 8;

QY 77 NIDTDQIIPAEHLTLVPSKPEDEYRKLGSFAFLPSAAYPTPFVAPGESSRY----- 129
DB 18 NVDTDQIIPKQFL-----KRIERTGYGRFA-----FDWKRIDANGEPNPEFELNOPY 65

QY 130 ---AIIVGANFGCGSSREHAPALGAAGARAIVAGYARIEFRNSVATGVEVPLELDV 186
DB 66 QGASIIIDAGENFGCGSSREHAPALDYGFKIITAPSFADIFHQNCFKNG-MLPIRM-py 123

QY 187 GAMKECKTGDV-----VTVDLANSVFINTSGKE 215
DB 124 DMWKQL-VGQYENOSIQMTVDLENQL-IHDSGNQ 156
```

Search completed: March 31, 2003, 09:08:09
Job time : 26 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 31, 2003, 09:05:48 ; Search time 33 Seconds

(without alignments)
1554.719 Million cell updates/sec

Title: US-10-027-450-47

Perfect score: 1266
Sequence: 1 MAAALSTAVSTALAPLR.....AGIFAVARKTGIMSKAAA 249

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPREMBL_21.*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mmc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_ricent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriaph:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	990.5	78.2	257	10	0949D1
2	710.5	56.1	251	10	092W85
3	651	51.4	253	10	091Y77
4	627	49.5	256	10	092W84
5	265.5	21.0	164	17	080280
6	260	20.5	163	16	097E81
7	258.5	20.4	164	17	090Z06
8	254.5	20.1	170	17	08TX94
9	252.5	19.9	161	16	08RDX1
10	252	19.9	162	17	08TU71
11	231	18.2	161	17	08ZM36
12	229.5	18.1	208	16	09RTT0
13	227.5	18.0	164	17	08TRF7
14	227.5	18.0	166	16	09WYC8
15	225.5	17.8	166	17	08TJM9
16	223.5	17.7	165	17	097YX3

17	222	17.5	168	17	0974Q9	0974Q9 sulfobolus
18	212	16.7	168	17	08TW31	08TW31 methanopyru
19	210	16.6	164	17	08U0B9	08U0B9 pyrococcus
20	206.5	16.3	163	17	09V1I9	09V1I9 pyrococcus
21	202	16.0	163	2	092ND9	092ND9 thermus the
22	192	15.2	659	16	067556	067556 aquifex aeo
23	189	14.9	202	16	08YX03	08YX03 anabaena sp
24	188	14.8	215	16	09PAX1	09PAX1 xylella fas
25	184	14.5	212	16	09HZ44	09HZ44 pseudomonas
26	182.5	14.4	216	16	08XXX4	08XXX4 ralsstonia s
27	179.5	14.2	711	3	09P3Y5	09P3Y5 yarrowia li
28	178	14.1	196	2	09AIM2	09AIM2 streptococc
29	176.5	13.9	201	16	092LAL	092LAL rhizobium m
30	174.5	13.8	193	16	08Y5R6	08Y5R6 listeria mo
31	174.5	13.8	194	16	09K8P1	09K8P1 bacillus ha
32	172.5	13.6	641	16	08RCF8	08RCF8 thermoaer
33	171.5	13.5	202	16	09A8N1	09A8N1 caulobacter
34	170	13.4	201	2	08RP98	08RP98 methylodact
35	170	13.4	213	16	09JZ16	09JZ16 neisseria m
36	169.5	13.4	193	16	09ZAZ5	09ZAZ5 listeria in
37	169	13.3	213	16	09J081	09J081 neisseria m
38	168.5	13.3	201	16	098E51	098E51 rhizobium l
39	168	13.3	200	16	09K800	09K800 vibrio chol
40	167.5	13.2	216	16	08UBR0	08UBR0 agrobacteri
41	163	12.9	201	16	09C3N8	09C3N8 pasteurella
42	161.5	12.8	177	16	09RTY5	09RTY5 deinococcus
43	159.5	12.6	182	2	0938C8	0938C8 mycobacteri
44	159	12.6	190	16	0995J2	0995J2 staphylococ
45	158.5	12.5	211	16	08YCW7	08YCW7 brucella me

ALIGNMENTS

RESULT 1	ID	Q949D1	PRELIMINARY;	PRT:	257 AA.
AC	Q949D1				
DT	01-DEC-2001	(TREMBLrel. 19, Created)			
DT	01-DEC-2001	(TREMBLrel. 19, Last sequence update)			
DT	01-JUN-2002	(TREMBLrel. 21, Last annotation update)			
DE	Hypothetical 26.4 kDa protein.				
GN	C740ERIPDM.				
OS	Oryza sativa (Rice).				
CC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;				
CC	Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;				
CC	Ehharitoidae; Oryzae; Oryza.				
OX	NCBI_TaxID=4530;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE=21329048; PubMed=11435398;				
RA	Mayer K., Murphy G., Tarchini R., Wambutt R., Volckaert G., Pohl T.,				
RA	Duesterhoeft A., Stiekema W., Entian K.D., Terry N., Lemke K.,				
RA	Haese D., Hall C.R., van Dodeeweerd A.M., Tingey S.V., Mewes H.M.,				
RA	Bayan M., Bancroft I.,				
RT	"Conservation of microstructure between a sequenced region of the				
RT	genome of rice and multiple segments of the genome of Arabidopsis				
RT	thaliana."				
RL	Genome Res. 11:1167-1174(2001).				
DR	EMBL; AJ307662; CAC39061.1; -				
DR	InterPro; IPR000573; Aconitase_C.				
DR	Pfam; PF00694; Aconitase_C; 1.				
KW	Hypothetical protein.				
SO	SEQUENCE 257 AA; 26443 MW; 7EC420ECF735FB66 CRC64;				

Query Match 78.2%; Score 990.5; DB 10; Length 257;

Best Local Similarity 76.7%; Pred. No. 4.6e-76;

Matches 197; Conservative 18; Mismatches 33; Indels 9; Gaps 2;

QY	1	MAAALSTAVSTA-----ALAPRAPTSAFLRBSOLCHRLHSKC---RAAGIYVA 51
DB	1	MAAALSTAVSTA-ADAVTVLAPCPPTSPRTFRRSRMVAALCRPAKCHHSRPLTAVAAA 60

RA Narusaka M., Nguyen M., Palm C.J., Sakurai T., Satou M., Seki M.,
 RA Shin P., Southwick A., Shinozaki K., Davis R.W., Ecker J.R.,
 RA Theologis A.;
 RT "Full length cDNA of gene At3g58990 (GI:15231608).";
 RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AL163527; CAB86927.1; -;
 DR EMBL; AY065326; AAL38807.1; -;
 DR InterPro: IPR000573; Aconitase_C.
 DR Pfam: PF00694; Aconitase_C; 1.
 SQ SEQUENCE 253 AA; 27208 MW; A99AB88C4656B0DA CRC64;

Query Match 51.4%; Score 651; DB 10; Length 253;
 Best Local Similarity 59.7%; Pred. NO. 2.6e-47;
 Matches 135; Conservative 29; Mismatches 36; Indels 6; Gaps 4;

QY 24 SAEIR-RSOLTCRLHSLKCRAGSIVPAALAAAAGSSPSSEVAFGECEVFGNDIDTDQ 82
 DB 29 SPFLQKLSASTIFENYKPLTSSATITRVAASSSDGSESTREPHGLCFVLKNDIDTDQ 88
 QY 83 IIPAEHLTVPSKDEYRGLGSFAFAGLPSAAYPTFPVAPGESRVAIIVGANFGCGS 142
 DB 89 IIPAEGLTLPSPEDREKLGSPALNGLPK-FYNERFVVPGBMKSYVILIGDNFGCGS 147
 QY 143 SREHAPVALGAAGARAIVAEGYARIFFRNSVATGEVPLELTDVGAMRECKTGDTVDL 202
 DB 148 SREHAPVCLGAAGAKAVAESYARIFFRNCVATGEIFPLE-SEVRICDECKTGDTVTEH 206
 QY 203 ---ANSVFINTHSKREYKLPICGDAGPVLEAGCIFAVARKGMTAS 245
 DB 207 KEDGSSLLINHTTRKEYKLPIDAGPVIDAGGIFRVAKKGMIPS 252

RESULT 4
 Q9ZM84 PRELIMINARY; PRT; 256 AA.

AC 09ZM84;
 DT 01-MAY-1999 (TREMBLrel. 10, Created)
 DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE 3-isopropylmalate dehydratase, small subunit (Putative
 AT2g43100.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC Eurosids II; Brassicales; Brassicaceae; Arabidopsids.
 NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-CV. COLUMBIA;
 RX MEDLINE-20083487; PubMed-10617197;
 RA Lin X., Kaul S., Rounsley S.D., Shea T.P., Beutlo M.-I., Town C.D.,
 RA Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblyum T.V.,
 RA Buel C.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H., Moffat K.S.,
 RA Cronin L.A., Shen M., Vanaken S.E., Umayam L., Tallon L.J., Gill J.E.,
 RA Adams M.D., Carrera A.J., Creasy T.H., Goodman H.M., Somerville C.R.,
 RA Copenhaver G.P., Preuss D., Nierman W.C., White O., Eisen J.A.,
 RA Salzberg S.L., Fraser C.M., Venter J.C.;
 RT "Sequence and analysis of chromosome 2 of the plant Arabidopsis
 thaliana.";
 RL Nature 402:761-768(1999).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-CV. COLUMBIA;
 RA Lin X.;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Yamada K., Banh J., Banno F., Chang E., Dale J.M., Goldsmith A.D.,
 RA Lee J.M., Onodera C.S., Quach H.L., Tang C., Toriumi M., Mu H.C.,
 RA Yamamura Y., Yu G., Yu S., Bowser L., Carninci P., Chen H., Cheuk R.,
 RA Hayashizaki Y., Ishida J., Jones T., Kamiya A., Karlin-Neumann G.,
 RA Kawai J., Kim C., Koesema E., Lam B., Lin J., Meyers M.C., Miranda M.,

RA Narusaka M., Nguyen M., Palm C.J., Sakurai T., Satou M., Seki M.,
 RA Shin P., Southwick A., Shinozaki K., Davis R.W., Ecker J.R.,
 RA Theologis A.;
 RT "Full length cDNA of gene At2g43100 (GI:15224222).";
 RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RA Cheuk R., Chen H., Kim C.J., Koesema E., Meyers M.C., Banh J.,
 RA Bowser L., Carninci P., Dale J.M., Goldsmith A.D., Hayashizaki Y.,
 RA Ishida J., Jiang P.X., Jones T., Kamiya A., Karlin-Neumann G.,
 RA Kawai J., Lam B., Lee J.M., Lin J., Liu S.X., Miranda M., Narusaka M.,
 RA Nguyen M., Onodera C.S., Palm C.J., Phan P.K., Quach H.L., Sakurai T.,
 RA Satou M., Seki M., Southwick A., Tang C.C., Toriumi M., Yamada K.,
 RA Yamamura Y., Yu G., Yu S., Shinozaki K., Davis R.W., Theologis A.,
 RA Ecker J.R.;
 RT "Arabidopsis cDNA clones.";
 RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
 RN [5]

RP SEQUENCE FROM N.A.
 RC STRAIN-CV. COLUMBIA;
 RA Lin X., Kaul S., Shea T.P., Fujii C.Y., Shen M., Vanaken S.E.,
 RA Barnstead M.E., Mason T.M., Bowman C.L., Ronning C.M., Beutlo M.-I.,
 RA Carrera A.J., Creasy T.H., Buel C.R., Town C.D., Nierman W.C.,
 RA Fraser C.M., Venter J.C.;
 RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AC004450; AAC64299.1; -;
 DR EMBL; AY063828; AAL36184.1; -;
 DR EMBL; AY060594; AAL31219.1; -;
 DR EMBL; AC006224; AAM15160.1; -;
 DR InterPro: IPR000573; Aconitase_C.
 DR Pfam: PF00694; Aconitase_C; 1.
 SQ SEQUENCE 256 AA; 27043 MW; 56845DC3F4F90A65 CRC64;

Query Match 49.5%; Score 627; DB 10; Length 256;
 Best Local Similarity 54.2%; Pred. NO. 2.8e-45;
 Matches 136; Conservative 38; Mismatches 65; Indels 12; Gaps 6;

QY 4 ALSTGAVSTALLIAPRAPTSFIRRSOLTCRLHSLKCRAGS---SIVPAAAAAAGS 59
 DB 11 ALPSCSTKTSLSIAFRSP---FLRFNGSTSLIPSSISITRSGNSPPIIRRAAASSDS 67
 QY 60 SSP-SSAVHGCFCFVVDNDIDTDQIIPAEHLTVPSKDEYRGLGSFAFALPSAAVPTP 118
 DB 68 NEALATWTHGICLVYKNDIDTDQIIPAGACTFPSNOEDENIAHLSGLP-FHKTR 126
 QY 119 FVAPGESRVAIIVGANFGCGSSREHAPVALGAAGARAIVAEGYARIFFRNSVATGEV 178
 DB 127 FIEPGENRSKYSIIIGENFGCGSSREHAPVCLGAAGAKAVAESYARIFFRNSVATGEV 186
 QY 179 YPELTDVGAMRECKTGDTVDLANS--VFINTHSKREYKLPICGDAGPVLEAGCIFAV 236
 DB 187 FPLE-SEVRICDECKTGDTVTELSDSGLLTNHTTGKNYKIKSGIDGAPVADGIFAY 245
 QY 237 ARKGMIAKSA 247
 DB 246 ARMGMIIPSLA 256

RESULT 5
 Q8U2A0 PRELIMINARY; PRT; 164 AA.
 AC 08U2A0;
 DT 01-JUN-2002 (TREMBLrel. 21, Created)
 DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE Putative 3-isopropylmalate dehydratase small subunit.
 GN PF0939.
 OS Pyrococcus furiosus.
 OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;
 OC Pyrococcus.
 NCBI_TaxID=2261;
 RN [1]
 RP SEQUENCE FROM N.A.

QY	66	VFEHCECVVDNIDTFOIILPAEHLTLVPSKPDREKLGSFAAGJPSAAYPPFPVAPGE	125
		: : : :	:
Db	4	VIRBAMVPFDDIDFDQIIIGRGLT--TQDPPE---LAKHWEG-ADPEFP-----EK	50
QY	126	SSRAITIVGANGFCGGSSREHPALGAAGAIAVEGARIFFRNSVATGEVPLELTLD	185
		: :	:
Db	51	VREGDVIYAARKNGPCCGSSREHPAILKACIVVRSEARIFYRNAIMLG--LLVACP	108
QY	186	VGAAKECKTDVVTVDLANSVFINHSGKEYKLPICGD-AGSVIEAGGFVAARRKG	241
		: : : :	: : : :
Db	109	GVDDAFEDDGLEIVNLRESYVKNLDTGETELEKPLPDPMRILEARAGLVELIKREG	164
RESULT 9			
O8RDKI			
ID	O8RDKI	PRELIMINARY:	PRT; 161 AA.
AC	O8RDKI:		
DT	01-JUN-2002 (TREMBLrel. 21,	Created)	
DT	01-JUN-2002 (TREMBLrel. 21,	Last sequence update)	
DT	01-JUN-2002 (TREMBLrel. 21,	Last annotation update)	
DE	3-Isopropylmalate dehydratase small subunit.		
GN	LEUD OR TTE0018.		
OS	Thermoanaerobacter tengcongensis.		
OC	Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;		
OC	Thermoanaerobacteriales; Thermoanaerobacteriaceae; Thermoanaerobacter		
OX	NCB1_Taxid=119072;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN-MBAT / JCM11007;		
RX	MEDLINE=21992816; PubMed=11997336;		
RA	Bao Q., Tian Y., Li W., Xu Z., Xuan Z., Hu S., Dong W., Yang J.,		
RA	Chen Y., Xue Y., Xu Y., Lai X., Huang L., Dong X., Ma Y., Ling L.,		
RA	Tan H., Chen R., Wang J., Yu Y., Yang H.;		
RT	"A complete sequence of T. tengcongensis genome.";		
RL	Genome Res. 12:689-700(2002).		
DR	EMBL: AEO12976; AAA2335.1; -.		
KW	Complete proteome.		
SQ	SEQUENCE 161 AA; 17888 MW; 4C4A2BCF9CC2B68F CRC64;		

RX MEDLINE:21929760; PubMed:11932238;
 RA Gajalaan J.E., Nusbaum C., Roy A., Endrizzi M.G., Macdonald P.,
 RA Faltz W., Calvo S., Engels R., Smlnov S., Atmoud D., Brown A.,
 RA Allen N., Naylor J., Strange-Thomann N., Deatellano K., Johnson R.,
 RA Linton L., McKean P., McKernan K., Talamas J., Tirrell A., Ye W.,
 RA Zimmer A., Barber R.D., Cann I., Graham D.E., Grahame D.A., Guss A.M.,
 RA Hedderlich R., Ingram-Smth C., Kueltnet H.C., Krzycki J.A.,
 RA Leigh J.A., Li W., Liu J., Mkhropadzhay B., Reeve J.N., Smth C.,
 RA Slinger T.A., Unayam L.A., White O., White R.H., de Macario E.C.,
 RA Ferry J.G., Jarrell K.F., Jing J., Macario A.J.L., Paulsen I.,
 RA Pilchett M., Sowers K.R., Swanson R.V., Zinder S.H., Lander E.,
 RA Metcalf W.W., Birren B.;
 RT "The genome of *Methanococcus acetivorans* reveals extensive metabolic
 RT and physiological diversity.";
 RL Genome Res. 12:532-542(2002)
 RL EMBL: AE010678; AAC03655.1; -.
 KW Complete proteome.
 QO SEQIDENCE 162 AA; 17625 MW; 2D609D9BA176a228 CRC64;

Query Match	15.9%	Score 252;	DB 17;	Length 162;
Best Local Similarity	35.8%	Pred. No. 9, 1e-14;		
Matches	62;	Conservative	31;	Mismatches 60; Indels 20; Gaps 7.
Qy	69	GECEVGDNDITDQIIIPAHLTVPSKPDPEYKLGSAFAGL-PSAAYPTPFVAPGSESS	127	
	: : : : : : : : : : : : : : : : : : : : : : : : :			
Db	3	GRAMKEGDVDVDAVIPGRY--LIENTPGE--LAKYFEEGVDDFA-----KKVH	48	
Qy	128	RYALIVGAGNFGCGSSREHAPALGAGARALVAGCYARIRFRNSVATGEYPLIELTDVG	187	
	: : : : : : : : : : : : : : : : : : : : : : : : :			
Db	49	ENDLIVAGSNFGSSREHAPALGKSKVCSVIAKSFARIFRNAINIG--VPVLECPNT--	106	
Qy	188	AMKECKTQGVVYVDLANSVFINTHSTSKREYKLKLPID--AGPIEAGGIFPAAR	239	
	: : : : : : : : : : : : : : : : : : : : : : : : :			
Db	107	--DRIDGDELEVDYSTGDIQNTITGETYQATPLDVEIRVETIDCGLETARR	157	

RESULT 11			
Q8ZW36			
ID	Q8ZW36	PRELIMINARY:	PRT; 161 AA.
AC	Q8ZW36;		
DT	01-MAR-2002 (TREMBLrel. 20, Created)		
DT	01-MAR-2002 (TREMBLrel. 20, Last sequence update)		
DT	01-JUN-2002 (TREMBLrel. 21, Last annotation update)		
DE	3-isopropylmalate dehydratase small subunit (leud).		
CN	PAE191.		
OS	Pyrobaculum aerophilum.		
OC	Archaea; Crenarchaeota; Thermoprotei; Thermoproteales;		
OC	Thermoproteaceae; Pyrobaculum.		
NCBI	NCBI_TaxID=13773;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	SRRAIN-IM2 / ATCC 51768 / DSM 7523;		
RX	PubMed=11792869;		
RA	Fitz-Gibbon S.T., Ladner H., Kim U.-J., Stetter K.O., Simon M.I.,		
RT	Miller J.H.;		
RT	"genome sequence of the hyperthermophilic crenarchaeon Pyrobaculum		
RT	aerophilum.";		
RL	Proc. Natl. Acad. Sci. U.S.A. 99:984-989(2002).		
DR	EMBL; AE009851; AAL63866.1; -		
DR	Interpro; IPR000573; Acontase_C.		
DR	Pfam; PF006594; Acontase_C; 1.		
CM	Complete proteome.		
SQ	SEQUENCE 161 AA; 17175 MW; 440B9AA6A3F4F5A2 CRC64;		

		Query Match	18.2%;	Score 231;	DB 17;	Length 161;	
		Best Local Similarity	36.4%;	Pred. No.	5.5E-12;		
		Matches	63;	Conservative	29;	Mismatches	63; Indels 18; Gaps 6;
QY	69	GECVYGDNDITDQTIIIPAEHLTVPSKPDPEYKLGSGFAFLGSPAAATPTPVNAGDESSR	128				
		: :					
Db	5	GRALVTGDKLIDTIVIIIPAKLVYT-----DPALLQDHAMEPL-DPEEP-----KKAK	50				
		: :					
YY	129	VAILVGANFGCGSSREHAPVALGAAGAARAVIAEGYAIRIFRNVSATGVGYPLELTIDVGA	188				

```

Db      51 GATVAGRAFGMGSSREQAALKAGVLAAYAESFARIFFRNAINVG--LPV-LQAPGI 107
QY      189 WKECKTGVYVDLANSVFINTSGEYKLPDI-GBAGVIAGGIFAVART 240
      108 REKKVKGDEVELDEGGIVNITTGGEIVKPLRGLPEILKAGGLNLYKNS 160

RESULT 12
O9RT10  PRELIMINARY;      PRT;      208 AA.
AC      O9RT10;
DT      01-MAY-2000 (TREMBLrel. 13, Created)
DT      01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DE      01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE      3-isopropylmalate dehydratase, small subunit.
GN      DR1784.
OS      Deinococcus radiodurans.
OC      Bacteria; Thermus/Deinococcus group; Deinococci; Deinococcales;
OC      Deinococcaceae; Deinococcus.
OX      NCBI_TaxID=1299;
RN      [1]
RC      SEQUENCE FROM N.A.
RA      MEDLINE=20036896; PubMed=10567266;
RA      White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,
RA      Dodson R.J., Qian D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,
RA      Moffat K.S., Haft H., Jiang L., Pamphile W., Crosby M., Shen M.,
RA      Vamatheva K.S., Lam P., McDonald L., Uterback T., Zalewski C.,
RA      Makarov K.A., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,
RA      Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,
RA      Fraser C.M.
RT      "Genome sequence of the radioresistant bacterium Deinococcus
RT      radiodurans R1."
RL      Science 286:1571-1577(1999).
DR      EMBL; AE002019; AAF11332.1; -.
DR      TIGR; DR1784; -.
DR      InterPro: IPR000573; Aconitase-C.
DR      Pfam; PF00694; Aconitase_C; 1.
KW      Complete proteome.
SQ      SEQUENCE 208 AA; 22652 MW; F26C4BB27FEBD1E7 CRC64;

Query Match      18.1%; Score 229.5; DB 16; Length 208;
Best Local Similarity 38.3%; Pred. No. 1e-11;
Matches 64; Conservative 21; Mismatches 59; Indels 23; Gaps 6;

QY      76 DNIDTDIIPAEHLTVPSKPEDEYRKLGSAFAAGLPSAAYPTPP---VAAGESSRAIT 132
      10 DHINTDEIIPARHLT-----TDVESELAKYAMED-----YDKDFVRRVQPD-----II 53
Db      133 VGSANFGGSSREHAPVALGAGARATVAGYARIFFRNSVATGEVYPLETIDVGAKREC 192
      54 VAGADFGGSSREHAWMLRGAGSAYIAPFAIIRYNSINNGF---LALECGITIELF 110
QY      193 KTGADVVDLANSVFINTSGEYKLPKIPG-AGPVIAGGIFAYAR 238
      111 QDGEAEHLDTKGITIRNRTKEISFVVPVQFALDVQKAGGMLYMK 157

RESULT 13
O9RT10  PRELIMINARY;      PRT;      164 AA.
AC      O9RT10;
DT      01-JUN-2002 (TREMBLrel. 21, Created)
DT      01-JUN-2002 (TREMBLrel. 21, Last sequence update)
DT      01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE      3-isopropylmalate dehydratase.
GN      LEUC OR MA1223.
OS      Methanosarcina acetivorans.
OC      Archaea; Euryarchaeota; Methanococci; Methanosarcinales;
OC      Methanosarcinaceae; Methanosarcina.
OX      NCBI_TaxID=2214;
RN      [1]

```

```

RP      SEQUENCE FROM N.A.
RC      STRAIN=C2A / ATCC 35395 / DSM 2834;
RX      MEDLINE=21929760; PubMed=11932238;
RA      Galagan J.E., Nusbaum C., Roy A., Endrizzi M.G., MacDonald P.,
RA      FilzHugh W., Calvo S., Engels R., Smirnov S., Atwood D., Brown A.,
RA      Allen N., Naylor J., Stange-Thomann N., DeArillano K., Johnson R.,
RA      Linton L., McEwan P., McKernan K., Talamas J., Tirrell A., Ye W.,
RA      Linton A., Barber R.D., Cann I., Graham D.E., Guss A.M.,
RA      Hedderich R., Ingram-Smith C., Kuetner H.C., Krzycki J.A.,
RA      Leigh J.A., Li W., Liu J., Mukhopadhyay B., Keefe J.N., Smith K.,
RA      Springer T.A., Umayam L.A., White O., White R.H., de Macario E.C.,
RA      Ferry J.G., Jarrell K.F., Jing H., Macario A.J.L., Paulsen I.,
RA      Pritchett M., Sowers K.R., Swanson R.V., Zinder S.H., Lander E.,
RA      Metcalf W.W., Birren B.;
RT      "The genome of Methanosarcina acetivorans reveals extensive metabolic
RT      and physiological diversity."
RL      Genome Res. 12:532-542(2002).
DR      EMBL; AE010790; AAM04642.1; -.
KW      Complete proteome.
SQ      SEQUENCE 164 AA; 17963 MW; 7D4660C68589AC0 CRC64;

Query Match      18.0%; Score 227.5; DB 17; Length 164;
Best Local Similarity 37.0%; Pred. No. 1.1e-11;
Matches 64; Conservative 25; Mismatches 61; Indels 23; Gaps 7;

QY      69 GECFVGNIDTDIIPAEHLTVPSKPEDEYRKLGSAFAAGLPSAAYPTPPVAFGESSR 128
      7 GWNKMFQGDINDTAIIIPKYL-----RTRDMQIFGTNAMEGIDPEF--TKRAKPGD---- 55
Db      129 VAIIVGANFGGSSREHAPVALGAGARATVAGYARIFFRNSVATGEVYPLETIDVGA 188
      56 --IIIVAGNFGGSSREDAPLAKSGIACIASFAIFFRNAINIG--LPLMEADV-- 109
QY      189 WKECKTGVYVDLANSVFINTSG--KEYKLPKIPGAGPVIAGGIFAYAR 239
      110 --ECQGEDEIKVLDKRGVIVPEKCIFFGNKLPDP--LLDILNDGGLVAHHKK 158
Db

RESULT 14
O9WYC8  PRELIMINARY;      PRT;      166 AA.
AC      O9WYC8;
DT      01-NOV-1999 (TREMBLrel. 12, Created)
DT      01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DT      01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE      3-isopropylmalate dehydratase, small subunit, putative.
GN      TM0292.
OS      Thermotoga maritima.
OC      Bacteria; Thermotogae; Thermotogaceae; Thermotoga.
OX      NCBI_TaxID=2336;
RN      [1]
RC      SEQUENCE FROM N.A.
RC      STRAIN=MSB / DSM 3109;
RX      MEDLINE=99287316; PubMed=10360571;
RA      Nelson K.E., Clayton R.A., Gill S.R., Gwinn M.L., Dodson R.J.,
RA      Haft D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,
RA      McDonald L., Uterback T.R., Malek U.A., Linher K.D., Garrett M.M.,
RA      Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.,
RA      Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,
RA      Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.;
RT      "Evidence for lateral gene transfer between Archaea and Bacteria from
RT      genome sequence of Thermotoga maritima."
RL      Nature 399:323-329(1999).
DR      EMBL; AE001711; MAD35380.1; -.
DR      TIGR; TM0292; -.
DR      InterPro: IPR000573; Aconitase-C.
DR      Pfam; PF00694; Aconitase_C; 1.
KW      Complete proteome.
SQ      SEQUENCE 166 AA; 18734 MW; DBACF34020511A4 CRC64;

Query Match      18.0%; Score 227.5; DB 16; Length 166;
Best Local Similarity 34.2%; Pred. No. 1.1e-11;
Matches 64; Conservative 26; Mismatches 60; Indels 37; Gaps 7;

```


Search completed: March 31, 2003, 09:08:53
Job time : 37 secs

```
OY 66 VFHGEFVVDNIDTDQIIPAEHLTLVPSKPDPEYRK-----LGSFAPAGLPSAAYPTPF 119
DB 1 MINGRAWKFGDNDISTDHIAPGRYFHL-RNNLELAKHVLDEAMEDFA----- 46
OY 120 VAPGESSRAYIIVGANGCGSSREHAPVALGAAGARAIYAEGARIFFRNSVATGEVY 179
DB 47 ----KKVQKGDILVAKNGKGLSSRREHARIITKINGVSCIYAKSFARIFFRNAIVNG-LP 101
OY 180 PLELTDVGAMKECKTGDVYTVDLANSVFTNHTSGKEYKLKPIGDAGP-----VIEAGIGF 234
DB 102 VIELKEV---DEINQDELEIDLENGVLKMLTKGKEYRFTPI---PKELLEILKEDGIV 154
OY 235 AYARKTG 241
DB 155 NYLKKHG 161
```

RESULT 15

```
O8TJM9 PRELIMINARY; PRT; 166 AA.
ID O8TJM9
AC O8TJM9;
DT 01-JUN-2002 (TREMblrel. 21, Created)
DT 01-JUN-2002 (TREMblrel. 21, Last sequence update)
DE 01-JUN-2002 (TREMblrel. 21, Last annotation update)
DE 3-Isopropylmalate dehydratase.
LEUD OR MA3751.
OS Methanosarcina acetivorans.
OC Archaea; Euryarchaeota; Methanococci; Methanosarcinales;
OX Methanosarcinaceae; Methanosarcina.
OX NCBI_TaxID=2214;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C2A / ATCC 35395 / DSM 2834;
RX MEDLINE=21929760; PubMed=11932238;
RA Galagan J.E., Nusbaum C., Roy A., Endrizzi M.G., Macdonald P.,
RA Fitzhugh W., Calvo S., Engels R., Smirnov S., Atnoor D., Brown A.,
RA Allen N., Naylor J., Stange-Thomann N., Dearellano K., Johnson R.,
RA Linton L., McKernan P., McKernan K., Talamas J., Tirrell A., Ye W.,
RA Zimmer A., Barber R.D., Cann I., Graham D.E., Grahame D.A., Guss A.M.,
RA Hedderich R., Ingram-Smith C., Kuetner H.C., Krzycki J.A.,
RA Leigh J.A., Li W., Liu J., Mukhopadhyay B., Reeve J.N., Smith K.,
RA Springer T.A., Umayam L.A., White O., White R.H., de Macario E.C.,
RA Ferry J.G., Jarrell K.F., Jing H., Macario A.J.L., Paulsen I.,
RA Pritchett M., Sowers K.R., Swanson R.V., Zinder S.H., Lander E.,
RA Metcalf M.W., Birren B.;
RT "The genome of Methanosarcina acetivorans reveals extensive metabolic
and physiological diversity.";
RL Genome Res. 12:532-542(2002).
DR EMBL; AE011086; AAM07104.1; -.
KW Complete proteome.
SQ SEQUENCE 166 AA; 18151 MW; 4BC4ED4738FDE014 CRC64;
```

Query Match 17.8%; Score 225.5; DB 17; Length 166;

Best Local Similarity 33.7%; Pred. No. 17e-11; Mismatches 49; Indels 45; Gaps 8;

```
DB 69 GECFVVDNIDTDQIIPAEHLTLVPSKPDPEYRKLSFAPAGLPSAAYPTPFVAPGESSR 128
OY 129 YAIIVGANGCGSSREHAPVALGAAGARAIYAEGARIFFRNSVATGEVYPLELTDVGA 188
DB 57 --IIVAGENGCGSSRQAPLAKHAGIACVYAKSFARIFFRNAIVNG--LELMADI-- 110
OY 189 WKECKTGDVYTVDL-----ANSVFINHTSGKEYKLKPIGDAGP-----VIEAGIGF 235
DB 111 --EQQEGDETEVDLLKGEVAVPQGVF-----VGNKLPDFLLDILITDGLYS 155
OY 236 YARK 239
DB 156 HRRK 159
```

